

CLUSTERANALYSE

auf topologisch-maßtheoretischer Grundlage

Paul O. Degens

1970



Digitized by the Internet Archive
in 2026

https://archive.org/details/IA41552314_0044

CLUSTERANALYSE

auf topologisch-maßtheoretischer Grundlage

Dissertation
zur Erlangung des akademischen Grades
eines Doktors der Naturwissenschaften
des Fachbereichs Mathematik
der Ludwig-Maximilians-Universität
München

vorgelegt von Paul O. Degens
aus Köln
im Mai 1978

CLUSTERANALYSE

mit besonderer Berücksichtigung der Grundlagen

Thema:

1. Referent: Professor Dr. U. Oppel

Korreferent: Professor Dr. H.G. Kellerer

Tag der mündlichen Prüfung: 6.10.1978

Ort: Leiden



1. Referent: Professor Dr. U. Oppel

Korreferent: Professor Dr. H.G. Kellerer

Tag der mündlichen Prüfung: 6.10.1978

Inhaltsverzeichnis

	Seite
Einleitung	4
Zusammenfassung	8
I. Teilmaße und schwache Topologie	13
II. Der Raum der Clusterungen	19
III. Gütefunktionen und Clusterungen	46
IV. Beispiele und Anwendungen	62
Anhang	75
Literaturverzeichnis	77
Lebenslauf	78

Einleitung

Ein wesentlicher Schritt bei jedem wissenschaftlichen Vorgehen ist die Klassifikation. So hat beispielsweise Linné 1758 die Lebewesen aufgrund ihrer Eigenschaften systematisch klassifiziert; seine Klassifikation war die Grundlage für die Evolutionstheorie und damit für den Fortschritt der Biologie selbst. Analog hat das Periodensystem als Klassifikation der Elemente für die Chemie, das Hertzsprung-Russell-Diagramm für die Klassifikation der Sterne in der Astrophysik große Bedeutung.

Mit Verfahren der Clusteranalyse versucht man, Objekte, die durch bestimmte Messungen beschrieben sind, in Gruppen einzuteilen, d.h. zu klassifizieren. Man kann beispielsweise die Sterne aufgrund ihrer scheinbaren Abstände (im Bogenmaß) so in Gruppen einteilen, daß jeweils nahe beieinanderliegende Sterne zu einer Gruppe gehören. Ling (1973) erhielt auf diese Weise mit einem einfachen Algorithmus wieder die bekannten Sternbilder zurück; es ist dabei interessant zu sehen, mit welchen formalen Algorithmen Strukturen gefunden werden können, die das menschliche Auge häufig wahrnimmt.

In der Literatur - eine recht umfassende Literaturübersicht enthält Bock (1974) - wird aufgrund spezieller heuristischer Überlegungen eine Vielzahl von Verfahren vorgeschlagen. Clusteranalytische Verfahren werden meist auf große Objektmengen angewandt, deren Struktur weitgehend unbekannt ist; diese Verfahren sollen dabei die "Methode des genauen Hinsehens" ersetzen. So ist es nicht verwunderlich, daß die Entwicklung von Algorithmen für die elektronischen

Datenverarbeitungsanlagen und hiermit zusammenhängende Probleme einen umfangreichen Teil der gesamten Literatur über Clusteranalyse darstellen. Der naheliegende und am meisten verwendete heuristische Ansatz, die Partition der Objekte als Clusterung zu wählen, die eine bestimmte Funktion (Gütefunktion, splitting function, objective function, Kriterium) maximiert bzw. minimiert, führt hierbei zu rechentechnischen Schwierigkeiten, da die Anzahl möglicher Partitionen mit wachsender Anzahl der Objekte extrem ansteigt. Man versucht dieses Problem zu umgehen, einmal indem man "hierarchisch agglomerativ" vorgeht, d.h. indem man aufgrund einer abgewandelten Gütefunktion sukzessive Partitionen durch Vereinigung von je zwei Klassen vergrößert, bis man ausgehend von der diskreten Partition mit einelementigen Klassen zur indiskreten Partition mit nur einer Klasse gelangt. Auf der anderen Seite kann man, von einer beliebigen Partition ausgehend, diese iterativ im Sinne der Gütefunktion verbessern - durch Austausch eines oder mehrerer Objekte (hill-climbing algorithm) -, bis man ein "lokales Maximum" erreicht. In gewissen Fällen kann man auch mit nicht-linearer Programmierung zum Ziele kommen, vgl. z.B. Gordon und Henderson (1977).

Ein anderer Problemkreis ergibt sich aus der statistischen Betrachtungsweise. Faßt man die Objektmenge als konkrete Stichprobe auf, dann kann man Clusterungen von Stichproben als Schätzungen der noch genauer zu definierenden Clusterung der zugrundeliegenden Verteilung ansehen. Soll diese Auffassung gerechtfertigt sein, so müssen die Clusterungen der Stichproben die Clusterung der zugrundeliegenden Verteilung immer besser approximieren, je größer der Stich-

probenumfang wird; denn man sucht ja Aussagen über die zugrundeliegende Verteilung und nicht über die zufällig herausgegriffene Stichprobe.

Es erweist sich daher als notwendig, topologische Strukturen einzuführen, die es beispielsweise erlauben von der Konvergenz der Clusterungen konkreter Stichproben gegen eine Clusterung der Grundgesamtheit zu sprechen. Eine solche rein asymptotische Betrachtung reicht aber für die praktischen Bedürfnisse meist nicht aus, besonders in den Situationen, in denen man Verfahren der Clusteranalyse anwendet: Wegen unzureichender Modellvorstellungen der jeweiligen Fachwissenschaft - diese erhofft man sich ja von den Ergebnissen einer Clusteranalyse - ist einerseits häufig nicht bekannt, welche Messungen besonders wichtig sind, andererseits werden fehlerhafte Messungen (insbesondere Ausreißer) kaum erkannt. Zusätzlich zur Konsistenz wird man daher gewisse Stetigkeitsbetrachtungen im Sinne der Theorie der Robustheit von statistischen Verfahren durchführen.

In der Literatur findet man über diese Probleme wenig, vgl. Bock (1974), der dies Problem höchstens anhand eines parametrischen Modells angeht. Jardine und Sibson (1971) halten einen statistischen Ansatz für nicht möglich. Rao (1972) schreibt: "The subject is still in its initial stages of development and the problems are not always clearly defined". Zu einem ähnlichen Urteil kamen schon Kendall und Stuart (1963): "The subject is far from being exhausted". Auch Huber (1975 a) hält eine grundlegende Klärung dieser Methoden für wichtig, aber ungelöst. Diese Arbeit soll einen Beitrag zur Lösung dieses statistischen Problems liefern.

An dieser Stelle möchte ich Herrn Prof. Ploog danken, daß ich diese theoretischen Untersuchungen neben meiner praktischen Tätigkeit am Max Planck Institut für Psychiatrie durchführen durfte. Herrn Hansert danke ich für viele nützliche Diskussionen, die zur Entwicklung der hier dargelegten Grundvorstellungen und des Modells führten, Herrn Hanisch und Herrn Oppel für Diskussionen und Hinweise bei der mathematischen Untersuchung.

Zusammenfassung

Statt ein Cluster als Teilmenge einer Objektmenge (meist eine konkrete Stichprobe) anzusehen, kann man Cluster als Teilmaße von Wahrscheinlichkeitsmaßen auffassen. Sei X ein separabler, metrisierbarer, topologischer Raum und $\mathcal{L}$ der zugehörige σ -Körper der Borelschen Teilmengen von X . Die Menge $\mathcal{M}$ aller nicht negativen, endlichen Maße auf $(X, \mathcal{L})$ sei mit der schwachen Topologie versehen. Ein Maß $\gamma \in \mathcal{M}$ heißt Teilmaß von $\mu \in \mathcal{M}$ ($\gamma \leq \mu$), wenn $\gamma(B) \leq \mu(B)$ für alle $B \in \mathcal{L}$ gilt. Man kann nun ein Maß $\mu \in \mathcal{M}^1$ in Teilmaße aufteilen; da aber nicht von vorneherein auszuschließen ist, daß einzelne Teilmaße in einer Aufteilung wiederholt vorkommen, betrachtet man auf eine höchstens abzählbare Indexmenge I bezogene Familien $(\mu_i)_I$ von Teilmaßen von μ mit $\sum_{i \in I} \mu_i = \mu$ oder allgemeiner mit $\sum_{i \in I} \mu_i \leq \mu$; sei dann

$$E_I^{\leq} := \{(\mu_i)_I \in \mathcal{M}^I : \sum \mu_i \leq \mu \in \mathcal{M}^1\} \quad \text{und}$$

$$E_I^1 := \{(\mu_i)_I \in \mathcal{M}^I : \sum \mu_i = \mu \in \mathcal{M}^1\}. \quad \sum_I \mu_i(X) \text{ heißt die Gesamtmasse von } (\mu_i)_I, \text{ sie ist auf } E_I^{\leq} \text{ höchstens gleich, auf } E_I^1 \text{ gleich 1.}$$

Da die Reihenfolge der Cluster (Teilmaße) μ_i von $(\mu_i)_I$ keine Rolle spielen soll und auch das Hinzufügen oder Weglassen von leeren Clustern (0-Maßen) keinen Einfluß haben soll, geht man zur entsprechenden Äquivalenzklasse $(\tilde{\mu}_i)_I$ über. $\tilde{E}_I^{\leq} := \{(\tilde{\mu}_i)_I : (\mu_i)_I \in E_I^{\leq}\}$ heißt der Raum der Clusterungen, der Teilraum $\tilde{E}_I^1 := \{(\tilde{\mu}_i)_I : (\mu_i)_I \in E_I^1\}$ heißt Raum der Clusterungen ohne Masseverlust.

Man kann nun von der schwachen Topologie auf $\mathcal{M}$ ausgehend zuerst auf $E_I^{\leq} \subset \mathcal{M}^I$ eine geeignete Topologie einführen und dann zur Quotiententopologie auf $\tilde{E}_I^{\leq}$ übergehen. Die Quotiententopologie läßt sich einfach charakterisieren (II Satz 3):

Clusterungen $(\tilde{\mu}_i^n)_I$ konvergieren genau dann gegen eine Clusterung $(\tilde{\mu}_i)_I$ für $n \rightarrow \infty$, wenn man Repräsentanten $(\mu_i^n)_I \in E_I^{\leq}$ von $(\tilde{\mu}_i^n)_I \in \tilde{E}_I^{\leq}$ wählen kann, die gegen einen Repräsentanten $(\mu_i)_I$ von $(\tilde{\mu}_i)_I$ für $n \rightarrow \infty$ konvergieren. Geht man von der Produkttopologie $\mathcal{T}$ auf $E_I^{\leq} \subset \mathcal{M}^I$ aus, so ist für endliches I oder bei Beschränkung auf $\tilde{E}_I^1$ die zugehörige Quotiententopologie $\tilde{\mathcal{T}}$ sogar auf natürliche Weise metrisierbar (II Satz 4 mit II Bem. 14, II Bem. 2). Für abzählbares I ist $\tilde{\mathcal{T}}$ auf $\tilde{E}_I^{\leq}$ jedoch zu grob, $\tilde{\mathcal{T}}$ erfüllt dann nicht einmal das Trennungsaxiom T_1 : Zu jeder beliebigen Clusterung $(\tilde{\mu}_i)_I \in \tilde{E}_I^{\leq}$ und zu jeder endlichen Teilmenge J von I gibt es Repräsentanten $(\mu_i)_I$ mit $\mu_i = 0$ für $i \in J$; also konvergiert die konstante Folge $((\tilde{\mu}_i)_I)_N$ von Clusterungen gegen $(\tilde{0})_I$ (II Satz 3).

Will man vermeiden, daß auf diese Weise "Masse verloren" geht, so kann man auf $E_I^{\leq}$ die Topologie $\mathcal{T}_{\infty}$ der gleichmäßigen Konvergenz wählen, die analog eine Quotiententopologie $\tilde{\mathcal{T}}_{\infty}$ auf $\tilde{E}_I^{\leq}$ induziert, die auf dem vollen Raum $\tilde{E}_I^{\leq}$ der Clusterungen metrisierbar ist (II Satz 4). Für endliches I stimmt die Topologie mit $\mathcal{T}$ bzw. $\tilde{\mathcal{T}}$ überein, für abzählbares I noch auf E_I^1 bzw. $\tilde{E}_I^1$, dem Raum der Clusterungen ohne Masseverlust. In II Bem. 6. c) wird eine spezielle Art von Masseverlust ausgeschlossen und dadurch die Topologie $\mathcal{T}_{\infty}$ bzw. $\tilde{\mathcal{T}}_{\infty}$ charakterisiert: $\mathcal{T}_{\infty}$ ist die grösste Topologie, die feiner als die Produkttopologie ist und diese spezielle Art von Masseverlust bei der Konvergenz nicht zuläßt.

Es liegt nahe, die Topologie $\mathcal{T}_1$ (und die Quotiententopologie $\tilde{\mathcal{T}}_1$) zu betrachten, die als grösste Topologie definiert werden kann, die feiner als $\mathcal{T}$ ist und in der die Gesamtmasse als Abbildung $\Sigma^X: E_I^{\leq} \rightarrow \mathbb{R}$, $(\mu_i)_I \mapsto \Sigma \mu_i(X)$ stetig ist. Auch $(E_I^{\leq}, \tilde{\mathcal{T}}_1)$ ist metrisierbar und stimmt bei endlichem

I auf dem vollen Raum $\tilde{E}_I^{\leq}$, bei abzählbarer Indexmenge I auf dem Raum $\tilde{E}_I^1$ der Clusterungen ohne Masseverlust mit $\tilde{\tau}$ und $\tilde{\tau}_\infty$ überein (II Satz 4 mit II Bem. 14, II Bem. 2).

Das Anliegen der Clusteranalyse besteht nun aber nicht in der Beschreibung aller möglichen Clusterungen, sondern es werden optimale Clusterungen von Maßen gesucht. Ordnet man jeder Clusterung $(\tilde{\mu}_i)_I$ von einem Maß μ eine reelle Güte $G(\mu, (\tilde{\mu}_i)_I)$ zu, so induziert die Ordnung von $\mathbb{R}$ eine Präordnung auf den Paaren $(\mu, (\tilde{\mu}_i)_I)$, wobei $(\tilde{\mu}_i)_I$ eine Clusterung von μ ist. Wird für alle Maße μ eines Definitionsbereiches $\mathcal{D} \subset \mathcal{M}^1$ eine größte Güte bei einer (und nur einer) Clusterung angenommen, so ist durch die Güte G implizit eine Abbildung ϕ auf dem Definitionsbereich $\mathcal{D}$ definiert, die jedem Maß μ die beste Clusterung (unter G) zuordnet. Im allgemeinen wird die optimale Güte jedoch manchmal nicht oder gleichzeitig bei mehreren Clusterungen angenommen; die Menge $\hat{\phi}_G(\mu) := \{(\tilde{\mu}_i)_I : G(\mu, (\tilde{\mu}_i)_I) \text{ ist maximal}\}$ enthält keine oder mehrere Elemente. Ist G jedoch stetig und die entscheidende Menge der Clusterungen kompakt, so ist $\hat{\phi}_G(\mu)$ nicht leer. Für die bevorzugte Wahl der Topologie $\tilde{\tau}_\infty$ auf $\tilde{E}_I^{\leq}$ ist nun entscheidend, daß die Menge der Clusterungen eines Maßes kompakt ist, also für stetiges G die Menge $\hat{\phi}_G(\mu)$ der optimalen Clusterungen von μ nicht leer ist. In einem geeigneten Sinne ist $\hat{\phi}_G(\mu)$ dann sogar stetig (auch hier ist eine Kompaktheitsaussage, die für $\tilde{\tau}_\infty$, aber nicht für $\tilde{\tau}_1$ gilt, entscheidend (II, Satz 5)), insbesondere ist eine durch G implizit definierte Funktion ϕ stetig, wenn sie durch G eindeutig definiert ist (III Satz 3). Zum Beweis dieses Satzes benötigt man noch eine weitere Eigenschaft, die für beide Topologien $\tilde{\tau}_\infty$ und $\tilde{\tau}_1$ auf $\tilde{E}_I^{\leq}$ erfüllt ist (II Satz 6). $\tilde{E}_I^{\leq}$ ist in sich genügend dicht, d.h.: Konvergiert eine Folge von Maßen $\mu^n \in \mathcal{M}^1$ gegen ein Maß $\mu \in \mathcal{M}^1$

und ist $(\tilde{\mu}_i)_I$ eine Clusterung von μ , so existieren Clusterungen $(\tilde{\mu}_i^n)_I$ von μ^n , die gegen $(\tilde{\mu}_i)_I$ konvergieren.

Bei diesen Überlegungen kann man sich auf geeignete (abgeschlossene und "in sich genügend dichte") Teilmengen der Clusterungen beschränken. Alle stetigen Clusterungsfunktionen, das sind stetige Funktionen, die Wahrscheinlichkeitsmaßen "ihre" Clusterung zuordnen, lassen sich durch Gütefunktionen implizit im obigen Sinne definieren (IIf Satz 4). Die Überlegungen, die die Beziehungen zwischen stetigen Gütefunktionen und Clusterungsfunktionen klarlegen, werden in vereinfachter Form im Anhang aufgeführt und liefern dort bei kompakter Parametermenge eine entsprechende Beziehung zwischen "Gütefunktionen" und Schätzfunktionen.

Die Clusterungen zu einer festen Indexmenge J kann man als abgeschlossenen Teilraum der Clusterungen zu einer abzählbaren Indexmenge I mit $|J| \leq |I|$ auffassen. Die Menge aller Clusterungen zu endlichen Indexmengen liegt im Raum der Clusterungen mit abzählbarer Indexmenge dicht. Daher wird man, wenn von einer Fragestellung her keine natürliche obere Schranke für die Anzahl der nicht leeren Cluster vorgegeben werden kann, im Raume der Clusterungen zu einer abzählbaren Indexmenge I arbeiten. (II Satz 7). Bezüglich der Topologie τ_∞ liegen die Clusterungen ohne Masseverlust und mit nur endlich vielen nicht leeren Clustern in $\tilde{E}_I^{\leq}$ dicht, für eine befriedigende Theorie der Clusteranalyse muß man daher Clusterungen mit Masseverlust mit betrachten.

In Kapitel IV (Satz 1) wird eine Klasse von bezüglich τ_∞ stetigen Gütefunktionen konstruiert, bei denen man leicht erzwingen kann, daß die zugehörigen optimalen Clusterungen in $\tilde{E}_I^1$ liegen. An einem Beispiel wird dann vorgeführt,

wie man unter etwas schwächeren Voraussetzungen noch die Konsistenz von Clustermethoden nachweisen kann. Einige Beispiele aus der Literatur zeigen, daß man den vorliegenden Ansatz in vielen Fällen sinnvoll auf vorhandene Verfahren anwenden kann.

Damit die Auffassung von einem Cluster als Teilmenge der Objektmenge im Rahmen der vorliegenden Arbeit miterfaßt wird, werden "mengenartige" Teilmaße, die solchen Teilmengen entsprechen, eingeführt und speziell behandelt. Die wesentliche Stetigkeitsaussage von III Satz 3 bleibt dabei als Konsistenzaussage (III Satz 3*) erhalten.

I Teilmaße und schwache Topologie

X sei ein separabler und metrisierbarer Raum, dessen Topologie von einer Metrik $d_X: X \times X \longrightarrow \mathbb{R}$ erzeugt wird. $C(X)$ sei die Menge aller stetigen, beschränkten, reellwertigen Funktionen auf X und $\mathcal{L}(X)$ die σ -Algebra der Borelschen Teilmengen von X . Die Menge $\mathcal{M}$ aller nicht-negativen endlichen Maße auf $\mathcal{L}(X)$ sei mit der schwachen Topologie $w(X)$ versehen. Mit X ist auch $\mathcal{M}$ ein metrisierbarer Raum mit abzählbarer Basis.

Definition

Es seien $\nu, \mu \in \mathcal{M}$. ν heißt Teilmaß von μ ($\nu \leq \mu$), wenn für jedes $B \in \mathcal{L}(X)$ gilt: $\nu(B) \leq \mu(B)$.

Offensichtlich ist dies gleichbedeutend damit, daß für alle $f \in C(X)$ mit $f \geq 0$ gilt: $\int f d\nu \leq \int f d\mu$.

Aus dieser Definition ergibt sich sofort

Satz 1 $(\nu^n)_{n \in \mathbb{N}}$ und $(\mu^n)_{n \in \mathbb{N}}$ seien gegen ν bzw. μ konvergente Folgen von Maßen. Ist für jedes $n \in \mathbb{N}$ das Maß ν^n ein Teilmaß von μ^n , so ist auch ν ein Teilmaß von μ .

Für eine Theorie der Clusteranalyse als Teilgebiet der Statistik sind empirische Maße, d.h. Maße μ mit $\mu = \frac{1}{N} \sum_{i=1}^N \delta_{x_i}$ besonders wichtig, wobei δ_x das Dirac-Maß in x ist.

Definition

Teilmaße ν eines empirischen Maßes $\mu = \frac{1}{N} \sum_{i=1}^N \delta_{x_i}$ mit $\nu = \frac{1}{N} \sum \text{gewisse } i \delta_{x_i}$ heißen mengenartig.

Dabei ist stillschweigend vorausgesetzt, daß der "Stichprobenumfang" N von μ fest gegeben ist. Bei einer Folge $(\mu^n)_{n \in \mathbb{N}}$ empirischer Maße mit $\mu^n \longrightarrow \mu$, nehmen wir dabei i.a. für die Stichprobenumfänge N_n von μ^n an, daß $N_n \longrightarrow \infty$ konvergiert.

Die Bezeichnung "mengenartig" soll die Auffassung von Clustern als "Teilmenge" von konkreten Stichproben hervorheben.

Einen weiteren für das Folgende wichtigen Zusammenhang zwischen den Begriffen Teilmaß und Konvergenz liefert

Satz 2

Es sei $(\mu^n)_{n \in \mathbb{N}}$ eine gegen $\mu \in \mathcal{M}$ konvergente Folge in $\mathcal{M}$.

ν sei ein Teilmaß von μ . Dann existiert eine gegen ν konvergente Folge $(\nu^n)_{n \in \mathbb{N}}$ von Teilmaßen mit $\nu^n \leq \mu^n$.

In speziellen Fällen kann man die Teilmaße ν^n noch besonders wählen.

1. Für empirische Maße μ^n mit $N_n \rightarrow \infty$ für $n \rightarrow \infty$ können die Teilmaße mengenartig gewählt werden.

2. Für Maße μ^n mit $\mu^n(X) \geq \mu(X)$ können die Teilmaße ν^n auf $\nu^n(X) = \nu(X)$ normiert werden.

Beweis:

1) Da X ein separabler und metrisierbarer Raum ist, existiert eine Folge $(\mathcal{Z}_k)_{k \in \mathbb{N}}$ von feiner werdenden, endlichen Zerlegungen $\mathcal{Z}_k$ mit $\mu(\partial Z) = 0$ für alle $Z \in \mathcal{Z}_k$, derart daß jede offene Teilmenge G von X als eine - von selbst abzählbare - Vereinigung von Elementen aus dem durchschnittstabilen System $\mathcal{Z} := \bigcup_{k \in \mathbb{N}} \mathcal{Z}_k$ darstellbar ist.

2) Für $n \in \mathbb{N}$ und $k \in \mathbb{N}$ seien die Maße $\eta_k^n: \mathcal{Z}(X) \rightarrow \mathbb{R}$ definiert durch

$$\eta_k^n(B) := \sum_{Z \in \mathcal{Z}_k} \frac{\nu(Z)}{\mu(Z)} \mu^n(B \cap Z),$$

dabei sei hier und im folgenden mit $\mu(Z) = 0$ auch $\frac{\nu(Z)}{\mu(Z)} = 0$.

Nach Konstruktion gilt $\eta_k^n \leq \mu^n$.

Für $Z \in \mathcal{Z}_{k^*}$ und $k \geq k^*$ gilt:

$$\eta_k^n(Z) = \sum_{Z' \in \mathcal{Z}_k} \frac{\nu(Z')}{\mu(Z')} \mu^n(Z' \cap Z) = \sum_{\substack{Z' \in \mathcal{Z} \\ Z' \in \mathcal{Z}_k}} \frac{\nu(Z')}{\mu(Z')} \mu^n(Z')$$

und

$$\nu(Z) = \sum_{\substack{Z' \in \mathcal{Z} \\ Z' \in \mathcal{Z}_k}} \nu(Z') = \sum_{\substack{Z' \in \mathcal{Z} \\ Z' \in \mathcal{Z}_k}} \frac{\nu(Z')}{\mu(Z')} \mu(Z').$$

Da $\mathcal{Z}_k$ endlich ist, gilt für

$$Z \in \mathcal{Z}_{k*} \quad \text{und} \quad k \geq k^* \quad \eta_k^n(Z) \longrightarrow \nu(Z).$$

3) Die Maße μ^n seien empirische Maße mit $N_n \longrightarrow \infty$ für $n \longrightarrow \infty$. Dann sei für $Z \in \mathcal{Z}, n \in \mathbb{N}$ definiert:

$$\lambda_{nZ} := \min(\lfloor \nu(Z) \cdot N_n \rfloor, \mu^n(Z) \cdot N_n);$$

dabei bezeichne $\lfloor \alpha \rfloor$ für $\alpha \in \mathbb{R}$ die größte ganze Zahl, die nicht größer als α ist. Weiter sei für $\mu^n = \frac{1}{N_n} \sum_{i=1}^{N_n} \delta_{x_{in}}$

$$\begin{aligned} I_{nZ} &:= \{i \in \{1, 2, \dots, N_n\} \mid x_{in} \in Z\} \\ &= \{i_1, i_2, \dots, i_{\lambda_{nZ}}\}. \end{aligned}$$

Die Anzahl λ_{nZ} der Elemente von I_{nZ} ist $\mu^n(Z) \cdot N_n$.

Wegen $\lambda_{nZ} \leq 1_{nZ}$ kann man aus I_{nZ} o.B.d.A. die ersten λ_{nZ} Elemente auswählen:

$$J_{nZ} := \{i_1, i_2, \dots, i_{\lambda_{nZ}}\}.$$

Für $n \in \mathbb{N}$ und $k \in \mathbb{N}$ seien die Maße $\tilde{\eta}_k^n: \mathcal{L}(X) \longrightarrow \mathbb{R}$ definiert:

$$\tilde{\eta}_k^n(B) := \frac{1}{N_n} \sum_{Z \in \mathcal{Z}_k} \sum_{i \in J_{nZ}} \delta_{x_{in}}(Z \cap B).$$

Nach Konstruktion gilt auch hier $\tilde{\eta}_k^n \leq \mu^n$.

Aus $\mu^n(Z) \longrightarrow \mu(Z)$ und $\frac{1}{N_n} \lfloor \nu(Z) N_n \rfloor \longrightarrow \nu(Z) \leq \mu(Z)$

für $n \longrightarrow \infty$ folgt

$$\frac{\lambda_{nZ}}{N_n} \longrightarrow \min(\nu(Z), \mu(Z)) = \nu(Z) \quad \text{für } n \longrightarrow \infty.$$

Für $Z \in \mathfrak{Z}_{k^*}$ und $k \geq k^*$ gilt:

$$\begin{aligned}\tilde{\eta}_k^n(Z) &= \frac{1}{N_n} \sum_{\substack{Z' \in \mathfrak{Z}_k \\ Z' \subseteq Z}} \sum_{i \in J_{nZ'}} \delta_{x_{in}}(Z' \cap Z) = \\ &= \frac{1}{N_n} \sum_{\substack{Z' \subseteq Z \\ Z' \in \mathfrak{Z}_k}} \sum_{i \in J_{nZ'}} \delta_{x_{in}}(Z') = \\ &= \frac{1}{N_n} \sum_{\substack{Z' \subseteq Z \\ Z' \in \mathfrak{Z}_k}} \lambda_{nZ'}.\end{aligned}$$

Da $\mathfrak{Z}_k$ für $k \in \mathbb{N}$ endlich ist, ergibt sich auch hier für

$Z \in \mathfrak{Z}_{k^*}$ und $k \geq k^*$ für $n \rightarrow \infty$:

$$\tilde{\eta}_k^n(Z) = \sum_{\substack{Z' \subseteq Z \\ Z' \in \mathfrak{Z}_k}} \frac{\lambda_{nZ'}}{N_n} \longrightarrow \sum_{\substack{Z' \subseteq Z \\ Z' \in \mathfrak{Z}_k}} \nu(Z') = \nu(Z).$$

4) Es sei $\mathfrak{Z}^k := \bigcup_{k' \leq k} \mathfrak{Z}_{k'}$.

In diesem Abschnitt ist im folgenden die Unterscheidung zwischen η_k^n und $\tilde{\eta}_k^n$ nicht erforderlich, da nur die in Abschnitt 2) bzw. 3) nachgewiesene Konvergenz ausgenutzt wird.

Für jede natürliche Zahl k und für alle $Z \in \mathfrak{Z}^k$ konvergiert

$$\eta_k^n(Z) \longrightarrow \nu(Z) \quad \text{für } n \rightarrow \infty.$$

Da wegen der Endlichkeit von $\mathfrak{Z}^k$ die Konvergenz auf $\mathfrak{Z}^k$ jeweils gleichmäßig ist, existiert zu jedem natürlichen k ein $n'_k \in \mathbb{N}$, so daß für alle $n \geq n'_k$ und für alle $Z \in \mathfrak{Z}^k$

$$|\eta_k^n(Z) - \nu(Z)| < \frac{1}{k} \quad \text{gilt.}$$

Zu jedem k sei nun ein solches $n'_k > k$ gewählt. Die Folge $(n_k)_{k \in \mathbb{N}}$ mit $n_k := \max(n'_1, n'_2, \dots, n'_k)$ konvergiert isoton gegen ∞ .

Sei $k(n) := k$ für $n_k \leq n < n_{k+1}$ ($k \geq 0$, $n_0 := 1$).

Dann ist nach Konstruktion $n \geq n_{k(n)}$, es gilt

$k(n) \rightarrow \infty$ für $n \rightarrow \infty$ und für $Z \in \mathfrak{Z}^{k(n)}$ und $k(n) \geq 1$:

$$\left| \eta_{k(n)}^n(Z) - \mu(Z) \right| < \frac{1}{k(n)}.$$

Für $Z \in \mathfrak{Z}$ gilt also $\eta_{k(n)}^n(Z) \rightarrow \mu(Z)$ für $n \rightarrow \infty$.

Da $\mathfrak{Z}$ ein schnittstabiles Mengensystem ist und sich jede offene Menge aus X als höchstens abzählbare Vereinigung von Elementen aus $\mathfrak{Z}$ darstellen läßt, ist $\mathfrak{Z}$ ein Konvergenz bestimmendes System, also konvergiert $\eta_{k(n)}^n =: \mu^n$ gegen μ . Hiermit ist die Hauptaussage des Satzes und Teil 1. bewiesen.

5) Gilt für die Maße zusätzlich $\mu^n(X) \geq \mu(X)$ für jedes $n \in \mathbb{N}$ so kann man nach 4) eine Folge μ^n wählen. O.B.d.A. sei auch $\mu(X) > 0$ vorausgesetzt, ebenso $\mu(X) > \mu(X)$.

Gilt $\mu(X) > \mu^n(X)$, so auch $\mu^n(X) \geq \mu(X) \geq \mu(X) > \mu^n(X)$ und damit $\mu^n(X) - \mu^n(X) > 0$. In diesem Fall kann man daher definieren:

$$\overline{\mu^n} := \mu^n + \frac{\mu(X) - \mu^n(X)}{\mu^n(X) - \mu^n(X)} (\mu^n - \mu^n).$$

Im anderen Falle gilt $\mu(X) \leq \mu^n(X)$; dies erlaubt die Definition:

$$\overline{\mu^n} := \frac{\mu(X)}{\mu^n(X)} \mu^n.$$

Die Maße $\overline{\mu^n}$ erfüllen nun nach Konstruktion die Bedingung

$$\overline{\mu^n}(X) = \mu(X) \quad \text{von 2.}$$

Q.e.d.

Im folgenden sei $\mathfrak{M}^1 := \{ \mu \in \mathfrak{M} : \mu(X) = 1 \}$ die Menge der Wahrscheinlichkeitsmaße auf $\mathfrak{L}(X)$ und $\mathfrak{M}^{\leq} := \{ \mu \in \mathfrak{M} : \mu(X) \leq 1 \}$. $\mathfrak{M}^{\leq}$ und $\mathfrak{M}^1$ sind in $\mathfrak{M}$ abgeschlossen.

Für $\mathcal{K} \subset \mathcal{M}^1$ sei $\mathcal{K}^{\leq} := \{ \mu \in \mathcal{M} : \mu \leq \mu \text{ für ein } \mu \in \mathcal{K} \}$.

Dann gilt folgender Satz:

Satz 3

Ist $\mathcal{K} \subset \mathcal{M}^1$ kompakt, so auch $\mathcal{K}^{\leq}$.

Beweis: 1. $\mathcal{K} \subset \mathcal{M}$ ist genau dann relativ kompakt, wenn für jede monoton fallende, gegen 0 punktweise konvergente Folge $(f_n)_{n \in \mathbb{N}}$ in $C(X)$ gilt:

$$\sup_{\mu \in \mathcal{K}} \int f_n d\mu \longrightarrow 0 \quad \text{für } n \longrightarrow \infty.$$

(Siehe Varadarajan (1966) II, Theorem 25)

Diese Bedingung gilt für $\mathcal{K}$ und damit erst recht für $\mathcal{K}^{\leq}$.

2. $\mathcal{K}^{\leq}$ ist abgeschlossen.

Sei nämlich $(\mu^n)_{n \in \mathbb{N}}$ eine gegen ein $\mu \in \mathcal{M}$ konvergente Folge in $\mathcal{K}^{\leq}$. Für $n \in \mathbb{N}$ sei $\mu^n \in \mathcal{K}$ ein gemäß der Definition von $\mathcal{K}^{\leq}$ zu μ^n gehöriges W-Maß. Wegen der Kompaktheit von $\mathcal{K}$ besitzt die Folge $(\mu^n)_{n \in \mathbb{N}}$ eine gegen ein $\mu \in \mathcal{K}$ konvergente Teilfolge. Mit Satz 1 folgt daraus, daß μ ein Teilmaß von μ ist, also in $\mathcal{K}^{\leq}$ liegt.

Q.e.d.

Aus der Abgeschlossenheit einer Menge $\mathcal{K} \subset \mathcal{M}^1$ in $\mathcal{M}^1$ folgt im allgemeinen nicht die Abgeschlossenheit von $\mathcal{K}^{\leq}$ in $\mathcal{M}^{\leq}$, wie folgendes Beispiel lehrt:

Beispiel: Es sei $X := \mathbb{R}$ mit der natürlichen Topologie,

$$\mathcal{K} := \left\{ \frac{1}{2} \delta_n + \frac{1}{2} \delta_{1/n} : n \in \mathbb{N} \right\}. \quad \text{Dann ist } \mathcal{K} \text{ in } \mathcal{M}^1$$

abgeschlossen, die Folge $(\frac{1}{2} \delta_{1/n})_{n \in \mathbb{N}}$ liegt in $\mathcal{K}^{\leq}$,

ihr Limes $\frac{1}{2} \delta_0 \in \mathcal{M}^{\leq}$ aber nicht.

II Der Raum der Clusterungen

Clusteranalytische Methoden teilen endliche Mengen in Cluster auf, d.h. sie ordnen endlichen Mengen Partitionen (Clusterungen) dieser Mengen zu. Faßt man die endlichen Mengen als konkrete Stichproben auf, so kann man entsprechend die zugehörigen empirischen Maße in mengenartige Teilmaße zerlegen. Es liegt dann auch nahe, beliebige Maße in Teilmaße zu zerlegen.

Endliche Mengen kann man nur in endlich viele nicht leere Mengen (Cluster) zerlegen; es gibt aber a priori keine natürliche obere Schranke für die Anzahl der Cluster, wenn man beliebige endliche Mengen zuläßt, insbesondere wenn man konkrete Stichproben wachsenden Umfangs betrachtet. Für eine allgemeine Formalisierung der Clusteranalyse wird man deshalb Aufteilungen eines Maßes in abzählbar unendlich viele Teilmaße zulassen.

In diesem Kapitel werden die obigen Überlegungen verallgemeinert und präzisiert, ein Raum der Clusterungen definiert und darauf topologische Eigenschaften untersucht.

d sei eine Metrik auf $\mathcal{M}$, die die schwache Topologie induziert mit

$$|\mu(X) - \nu(X)| \leq d(\mu, \nu) \leq \max(\mu(X), \nu(X)) ;$$

die Existenz solcher Metriken ist gesichert, vgl. Kellerer (1972).

I sei eine endliche oder abzählbare Indexmenge, o.B.d.A. sei dabei $I \subset \mathbb{N}$; soweit nicht anders vermerkt, sei sogar stets $I = \mathbb{N}$ oder $I = \{1, 2, \dots, n\}$ vorausgesetzt.

Definition

$$E_I^b := \{ (\mu_i)_I \in \mathcal{M}^I : \sum_{i \in I} \mu_i(X) < \infty \}$$

$$E_I^{\leq} := \{(\mu_i)_I \in \mathcal{M}^I : \sum_{i \in I} \mu_i(X) \leq 1\}$$

$$E_I^1 := \{(\mu_i)_I \in \mathcal{M}^I : \sum_{i \in I} \mu_i(X) = 1\} .$$

Die Abbildung $\Sigma^X: E_I^b \longrightarrow \mathbb{R}$ sei definiert durch

$$(\mu_i)_I \longmapsto \Sigma_I \mu_i(X) ; \text{ sie ordnet jedem Element } (\mu_i)_I \in E_I^b$$

die "Gesamtmasse" zu. Weiter sei $\Sigma: E_I^b \longrightarrow \mathcal{M}$ durch

$$(\mu_i)_I \longmapsto \Sigma_I \mu_i \text{ gegeben.}$$

Auf E_I^b seien weiter drei Metriken definiert:

$$D((\mu_i)_I, (\nu_i)_I) := \sum_{i \in I} 2^{-i} \min(1, d(\mu_i, \nu_i))$$

$$D_{\infty}((\mu_i)_I, (\nu_i)_I) := \max_{i \in I} d(\mu_i, \nu_i)$$

$$D_1((\mu_i)_I, (\nu_i)_I) := \sum_{i \in I} d(\mu_i, \nu_i) .$$

τ , τ_{∞} bzw. τ_1 seien die von D , D_{∞} bzw. D_1 erzeugten Topologien auf E_I^b .

Bemerkungen

1. Elementar kann man für D , D_{∞} und D_1 die Eigenschaften einer Metrik nachweisen. Es gilt $D \leq D_{\infty} \leq D_1$; wegen $d(\mu, \nu) \leq \max(\mu(X), \nu(X))$ ist D_1 auf E_I^b endlich.

Im Falle $I = \mathbb{N}$ gilt für $(\mu_i)_I \in E_I^b$ und $(\nu_i)_I \in E_I^b$
 $d(\mu_i, \nu_i) \leq \max(\mu_i(X), \nu_i(X)) \longrightarrow 0$ für $i \longrightarrow \infty$;
das Maximum $\max_{i \in I} d(\mu_i, \nu_i)$ wird also für ein $i \in I$ angenommen. Aus der Definition folgt weiter sofort $D \leq 1$.

2. Es gilt: $\tau \subset \tau_{\infty} \subset \tau_1$,
für endliches I sogar: $\tau = \tau_{\infty} = \tau_1$.

Ist $|I| = \infty$, so gilt $\tau \subsetneq \tau_{\infty} \subsetneq \tau_1$, vgl. Beispiel 1 und 2 unten.

3. Bezüglich der drei Topologien τ , τ_{∞} , τ_1 ist die

Multiplikation mit nicht negativen Skalaren

$[0, \infty) \times E_I^b \longrightarrow E_I^b$, $(\alpha, (\mu_i)_I) \longmapsto (\alpha \mu_i)_I$ jeweils stetig.

4. Σ^X ist bzgl. $\mathcal{F}_1$ stetig, bzgl. $\mathcal{F}$ (und erst recht $\mathcal{F}_\infty$) nach unten halbstetig ; Gesamtmasse kann also nur verlorengehen.

Beweis zu 4.: Aus $|\Sigma_I \mu_i(X) - \Sigma_I \nu_i(X)| \leq \Sigma_I |\mu_i(X) - \nu_i(X)| \leq \Sigma_I d(\mu_i, \nu_i) = D_1((\mu_i)_I, (\nu_i)_I)$ folgt die gleichmäßige Stetigkeit von Σ^X bzgl. D_1 , also auch die Stetigkeit von Σ^X bzgl. $\mathcal{F}_1$.

Aus der Stetigkeit der Funktionen

$$(\mu_i)_I \longmapsto \Sigma_{i=1}^n \mu_i(X) \quad \text{für } n \in \mathbb{N}$$

folgt die Halbstetigkeit nach unten von Σ^X :

$$(\mu_i)_I \longmapsto \sup_{n \in \mathbb{N}} \Sigma_{i=1}^n \mu_i(X) = \Sigma^X(\mu_i)_I .$$

5. Aus Bemerkung 4 folgt sofort: $E_I^{\leq}$ ist bzgl. $\mathcal{F}$ (und damit auch bzgl. $\mathcal{F}_\infty$ und $\mathcal{F}_1$) abgeschlossen in E_I^b . Die Menge E_I^1 ist bzgl. $\mathcal{F}_1$ in E_I^b und $E_I^{\leq}$ abgeschlossen; jedoch für $|I| = \infty$ nicht bzgl. $\mathcal{F}_\infty$ (und $\mathcal{F}$) , vgl. hierzu Beispiel 2. Es gilt weiter für $(\mu_i)_I, (\nu_i)_I \in E_I^{\leq}$:

$$D((\mu_i)_I, (\nu_i)_I) \leq 1 \quad , \quad D_\infty((\mu_i)_I, (\nu_i)_I) \leq 1 \quad \text{und}$$

$$D_1((\mu_i)_I, (\nu_i)_I) \leq 2 .$$

Beispiele für $\tau \neq \tau_\infty \neq \tau_1$.

Es sei $I = \mathbb{N}$, $0 \neq \mu \in \mathcal{M}^{\leq}$ und $(\mu_i)_I := (0, 0, 0, \dots)$.

1. Setze

$$\mu_i^n := \begin{cases} \mu & \text{für } i=n \\ 0 & \text{sonst.} \end{cases}$$

Dann gilt:

$D_\infty((\mu_i^n)_I, (\mu_i)_I) = d(\mu, 0) = \mu(X)$ für jedes $n \in \mathbb{N}$, aber

$D((\mu_i^n)_I, (\mu_i)_I) = 2^{-n} \mu(X)$.

$(\mu_i^n)_I$ konvergiert also bzgl. τ , nicht jedoch bzgl. τ_∞ gegen $(\mu_i)_I$.

2. Setze

$$\mu_i^n := \begin{cases} n^{-1} \mu & \text{für } i \leq n \\ 0 & \text{sonst.} \end{cases}$$

Dann gilt:

$D_1((\mu_i^n)_I, (\mu_i)_I) = \sum_{i \leq n} d(n^{-1} \mu, 0) = \mu(X)$ für jedes $n \in \mathbb{N}$.

Aber

$D_\infty((\mu_i^n)_I, (\mu_i)_I) = d(n^{-1} \mu, 0) = n^{-1} \mu(X)$.

$(\mu_i^n)_I$ konvergiert also bzgl. τ_∞ , nicht jedoch bzgl. τ_1 gegen $(\mu_i)_I$.

Satz 1

a) Sei $((\mu_i^n)_I)_{n \in \mathbb{N}}$ eine Folge in E_I^b und $(\mu_i)_I \in E_I^b$.

Dann konvergiert $(\mu_i^n)_I$ bzgl. τ_1 gegen $(\mu_i)_I$ genau dann, wenn $(\mu_i^n)_I$ bzgl. τ gegen $(\mu_i)_I$ konvergiert und zusätzlich die Gesamtmasse konvergiert.

b) In $E_I^{\leq}$ stimmen die Umgebungssysteme der Punkte aus E_I^1 bzgl. der drei Topologien τ , τ_∞ und τ_1 überein. Insbesondere haben diese Topologien auf E_I^1 dieselbe Spur.

Beweis:

Für endliches I folgt die Behauptung sofort aus Bemerkung 2 ; sei also o.E. $I = \mathbb{N}$.

a) " $\Rightarrow$ " folgt aus $\mathcal{T} \subset \mathcal{T}_1$ und Bemerkung 4.

" $\Leftarrow$ ": Wegen $(\mu_i)_{i \in E_I^b}$ existiert zu beliebigem $\varepsilon > 0$ ein $i_0 \in I$ mit $\sum_{i > i_0} \mu_i(X) < \varepsilon/4$. Aus der Konvergenz bzgl. $\mathcal{T}$

folgt $\sum_{i \leq i_0} \mu_i^n(X) \rightarrow \sum_{i \leq i_0} \mu_i(X)$ und unter Beachtung der Voraussetzung $\sum_I \mu_i^n(X) \rightarrow \sum_I \mu_i(X)$ für $n \rightarrow \infty$ auch

$$\sum_{i > i_0} \mu_i^n(X) = \sum_I \mu_i^n(X) - \sum_{i \leq i_0} \mu_i^n(X) \longrightarrow$$

$$\sum_I \mu_i(X) - \sum_{i \leq i_0} \mu_i(X) = \sum_{i > i_0} \mu_i(X).$$

Also gilt für beliebiges $\varepsilon > 0$, zugehöriges $i_0 \in I$ und genügend großes n

$$\begin{aligned} D_1((\mu_i^n)_I, (\mu_i)_I) &= \sum_I d(\mu_i^n, \mu_i) \leq \\ &\leq \sum_{i \leq i_0} d(\mu_i^n, \mu_i) + \sum_{i > i_0} \mu_i^n(X) + \sum_{i > i_0} \mu_i(X) \leq \\ &\leq \sum_{i \leq i_0} d(\mu_i^n, \mu_i) + 3\varepsilon/4. \end{aligned}$$

Aus $(\mu_i^n)_I \rightarrow (\mu_i)_I$ bzgl. $\mathcal{T}$ folgt $\sum_{i \leq i_0} d(\mu_i^n, \mu_i) \rightarrow 0$,

also gilt für beliebiges $\varepsilon > 0$ und genügend großes $n \in \mathbb{N}$

$D_1((\mu_i^n)_I, (\mu_i)_I) \leq \varepsilon$. Da $\varepsilon > 0$ beliebig war, folgt die Behauptung a).

b) Wegen Bemerkung 4 gilt für eine bzgl. $\mathcal{T}$ gegen $(\mu_i)_{i \in E_I^1}$ konvergente Folge $((\mu_i^n)_I)_N$ in $E_I^{\leq}$:

$$1 \geq \limsup_{n \rightarrow \infty} \sum_I \mu_i^n(X) \geq \liminf_{n \rightarrow \infty} \sum_I \mu_i^n(X) \geq \sum_I \mu_i(X) = 1.$$

Also konvergiert die Gesamtmasse $\sum_I \mu_i^n(X)$ gegen $\sum_I \mu_i(X)$.

Aus a) folgt nun $(\mu_i^n)_I \rightarrow (\mu_i)_I$ bzgl. $\mathcal{T}_1$.

Diese Überlegungen gelten für alle bzgl. $\mathcal{T}$ gegen ein $(\mu_i)_{i \in E_I^1}$ konvergenten Folgen in $E_I^{\leq}$. Da für die Topologien $\mathcal{T}$, $\mathcal{T}_\infty$ und $\mathcal{T}_1$

das erste Abzählbarkeitsaxiom gilt, folgt bei Beachtung von

$\mathcal{T} \subset \mathcal{T}_\infty \subset \mathcal{T}_1$ die Behauptung b).

Q.e.d.

Bemerkung

6. Die Topologien τ , τ_∞ und τ_1 hängen aufgrund der Definition nicht nur von der schwachen Topologie $w(X)$ auf $\mathcal{M}$, sondern von der speziellen Metrik d und der zugehörigen uniformen Struktur auf $\mathcal{M}$ ab. Diese Abhängigkeit ist aber im wesentlichen nur formaler Natur, wie die folgenden Überlegungen zeigen.

a) Die Topologie τ ist die Spur auf E_I^b der Produkttopologie auf $\mathcal{M}^I$ zu $w(X)$ auf $\mathcal{M}$.

b) Satz 1, a) liefert eine Charakterisierung der Topologie τ_1 ohne Benutzung der Metrik d .

c) Analog zu Satz 1, a) gibt es für die Topologie τ_∞ die folgende Charakterisierung ohne Benutzung der Metrik d :

Eine Folge $((\mu_i^n)_I)_{\mathbb{N}}$ konvergiert in E_I^b genau dann bzgl. τ_∞ gegen $(\mu_i)_I \in E_I^b$, wenn $((\mu_i^n)_I)_{\mathbb{N}}$ bzgl. τ gegen $(\mu_i)_I$ konvergiert und zusätzlich $\mu_i^n(X) \rightarrow 0$ für $(i, n) \rightarrow (\infty, \infty)$ gilt.

Beweis: " $\Rightarrow$ ": Es gilt:

$$\begin{aligned}\mu_i^n(X) &= d(\mu_i^n, 0) \leq d(\mu_i^n, \mu_1) + d(\mu_1, 0) \leq \\ &\leq D_\infty((\mu_i^n)_I, (\mu_1)_I) + \mu_1(X) .\end{aligned}$$

Unabhängig von i gilt für $n \rightarrow \infty$: $D_\infty((\mu_i^n)_I, (\mu_1)_I) \rightarrow 0$;

unabhängig von n gilt für $i \rightarrow \infty$: $\mu_i(X) \rightarrow 0$.

" $\Leftarrow$ ": Es gilt:

$$\begin{aligned}D_\infty((\mu_1^n)_I, (\mu_1)_I) &= \sup_I d(\mu_1^n, \mu_1) \leq \\ &\leq \max(\max_{i \leq i_0} d(\mu_1^n, \mu_1), \sup_{i > i_0} d(\mu_1^n, \mu_1)) \leq \\ &\leq \max(\max_{i \leq i_0} d(\mu_1^n, \mu_1), \sup_{i > i_0} \mu_1^n(X), \sup_{i > i_0} \mu_1(X)) .\end{aligned}$$

Zu beliebigem $\varepsilon > 0$ existiert wegen $\mu_1^n(X) \rightarrow 0$ für

$(i, n) \rightarrow (\infty, \infty)$ ein $i_0 \in I$ und ein $n_0 \in \mathbb{N}$ mit $\mu_i^n(X) < \varepsilon$

für $i \geq i_0$, $n \geq n_0$; wegen $(\mu_i)_I \in E_I^b$ existiert ein $i'_0 \in I$ mit $\mu_{i'_0}(X) < \varepsilon$ für $i \geq i'_0$ und wegen $(\mu_1^n)_I \longrightarrow (\mu_1)_I$ bzgl. τ existiert ein $n'_0 \in \mathbb{N}$ mit $d(\mu_{i'_0}^n, \mu_1) < \varepsilon$ für $n \geq n'_0$ und $i \leq \max(i_0, i'_0)$. Zusammen gilt für genügend großes n also: $D_\infty((\mu_1^n)_I, (\mu_1)_I) < \varepsilon$; da $\varepsilon > 0$ beliebig war, folgt die Behauptung.

Definition

Es sei $\mathcal{K} \subset \mathcal{M}^1$.

$$\begin{aligned} E(\mathcal{K}) &:= E_I^{\leq}(\mathcal{K}) := \left\{ (\mu_i)_I \in E_I^{\leq} : \sum_{i \in I} \mu_i \in \mathcal{K}^{\leq} \right\} \\ &= \left\{ (\mu_i)_I \in E_I^{\leq} : \sum_I \mu_i \leq \mu \text{ für ein } \mu \in \mathcal{K} \right\}. \end{aligned}$$

Satz 2

Ist $\mathcal{K}$ kompakt, so auch $E(\mathcal{K})$ bzgl. der Topologie τ .

Beweis:

1. $E_I^{\leq}$ ist in $\mathcal{M}^I$ abgeschlossen.

2. $E(\mathcal{K})$ ist abgeschlossen in $E_I^{\leq}$ bzgl. τ .

Gegeben sei eine Folge $((\mu_i^n)_I)_{n \in \mathbb{N}}$ in $E(\mathcal{K})$ und ein $(\mu_i)_I \in E_I^{\leq}$ mit $(\mu_i^n)_I \longrightarrow (\mu_i)_I$ bzgl. τ für $n \longrightarrow \infty$. Dann gilt für alle $n \in \mathbb{N}$ mit jeweils einem geeigneten $\mu^n \in \mathcal{K}$:

$\sum_{i \in I} \mu_i^n \leq \mu^n$. Da $\mathcal{K}$ kompakt ist, existiert eine gegen ein $\mu \in \mathcal{K}$ konvergente Teilfolge $(\mu^k)_{k \in \mathbb{N}}$ der Folge $(\mu^n)_{n \in \mathbb{N}}$.

Wäre nun für ein $B \in \mathcal{B}(X)$ etwa $\mu(B) < \sum_{i \in I} \mu_i(B)$, so gäbe es ein $j \in I$ mit $\mu(B) < \sum_{i \leq j} \mu_i(B)$.

I, Satz 1 angewandt auf $\mu^k := \sum_{i \leq j} \mu_i^k, \mu^k, \mu := \sum_{i \leq j} \mu_i$ und μ ergibt aber $\sum_{i \leq j} \mu_i \leq \mu$. Dies wäre ein Widerspruch, es gilt also $\sum_I \mu_i \leq \mu$, d.h. $(\mu_i)_I \in E(\mathcal{K})$.

3. Nach I, Satz 3 und dem Satz von Tychonoff ist $(\mathcal{K}^{\leq})^I$ in $(\mathcal{M}, w(X))^I$ kompakt. $E(\mathcal{K})$ ist wegen Teil 1 und Teil 2 in $(\mathcal{M}, w(X))^I$ abgeschlossen, als Teilmenge von $(\mathcal{K}^{\leq})^I$ also kompakt in $(\mathcal{M}, w(X))^I$ und damit auch in $E_I^{\leq}$ bzgl. $\mathcal{F}$.

Q.e.d.

Unter einer Clusterung soll eine Zerlegung eines W-Maßes in Teilmaße verstanden werden, dabei soll eine Reihenfolge der Teilmaße aber keine Rolle spielen. Für die entsprechenden Überlegungen benötigen wir einige Hilfsmittel:

Definition

$$\begin{aligned}\mathcal{I}(I) &:= \{ \sigma: I \longrightarrow I : \sigma \text{ ist bijektive Abbildung} \} \\ \mathcal{I}_0(I) &:= \{ \sigma \in \mathcal{I}(I) : \sigma i = i \text{ für alle bis auf endlich viele } i \in I \}\end{aligned}$$

Durch die Komposition

$$\mathcal{I}(I) \times E_I^b \longrightarrow E_I^b, \quad (\sigma, (\mu_i)_I) \longmapsto (\mu_{\sigma i})_I$$

operiert $\mathcal{I}(I)$ auf E_I^b .

Bemerkungen

7. Jedes $\sigma \in \mathcal{I}(I)$ liefert eine topologische Abbildung $\sigma: E_I^b \rightarrow E_I^b$, $(\mu_i)_I \mapsto (\mu_{\sigma i})_I = \sigma(\mu_i)_I$ bzgl. der Topologien $\mathcal{F}$, $\mathcal{F}_\infty$ und $\mathcal{F}_1$. Bezüglich der Metriken D_∞ und D_1 ist $\sigma \in \mathcal{I}(I)$ sogar eine Isometrie, offensichtlich gilt dies nicht für D . Die Räume $E_I^{\leq}$ und E_I^1 sind invariant unter $\sigma \in \mathcal{I}(I)$.

8. $\mathcal{I}_0(I)$ ist die Untergruppe von $\mathcal{I}(I)$, die durch die Transpositionen erzeugt wird. Identifiziert man für $J \in I$ $\sigma \in \mathcal{I}(I)$ mit $\sigma' \in \mathcal{I}(J)$, falls $\sigma i = \sigma' i$ für $i \in J$ und $\sigma i = i$ sonst,

so gilt $\mathcal{T}_0(I) = \bigcup_{J \subseteq I, J \text{ endl.}} \mathcal{T}(J)$ und speziell

$$\mathcal{T}_0(N) = \bigcup_{n \in \mathbb{N}} \mathcal{T}(\{1, 2, \dots, n\}) .$$

Die Permutationen $\sigma \in \mathcal{T}_0(I)$ liefern bzgl. der Metrik D in beiden Richtungen gleichmäßig stetige Abbildungen

$$\sigma: E_I^b \longrightarrow E_I^b .$$

Auf E_I^b kann man die folgenden Äquivalenzrelationen einführen:

Definition

$$(\mu_i)_I \approx (\nu_i)_I \quad : \Leftrightarrow \text{Es existiert } \sigma \in \mathcal{T}_0(I) \text{ mit } \sigma(\mu_i)_I = (\nu_i)_I .$$

Für ein $(\mu_i)_I \in E_I^b$ sei $I((\mu_i)_I) := I_\mu := \{i \in I : \mu_i \neq 0\}$,

weiter sei $I((\nu_i)_I) = I_\nu$, dann sei definiert:

$$(\mu_i)_I \sim (\nu_i)_I \quad : \Leftrightarrow \text{Es existiert } \sigma: I_\nu \longrightarrow I_\mu, \text{ eine bijektive Abbildung, mit } \nu_i = \mu_{\sigma i} \text{ für } i \in I_\nu .$$

Die zugehörigen Äquivalenzklassen von $(\mu_i)_I$ bezeichnen wir mit $\pi_0(\mu_i)_I$ bzw. $\pi(\mu_i)_I$. Statt $\pi(\mu_i)_I$ schreiben wir auch $(\tilde{\mu}_i)_I$.

Bemerkungen

2. a) $\approx$ und $\sim$ sind Äquivalenzrelationen mit

$\approx \subset \sim$ als Teilmengen von $E_I^b \times E_I^b$. Für $(\mu_i)_I \in E_I^b$ gilt $\pi_0(\mu_i)_I \subset \pi(\mu_i)_I$.

b) Für endliches I stimmen die beiden Äquivalenzrelationen überein.

c) Die Abbildungen $\Sigma: E_I^b \longrightarrow \mathcal{M}$, $(\mu_i)_I \longmapsto \sum_{i \in I} \mu_i$ und Σ^X sind auf den Äquivalenzklassen konstant.

$$d) \quad \pi_0(\mu_i)_I = \bigcup_{\sigma \in \mathcal{T}_0(I)} \{\sigma(\mu_i)_I\} .$$

Setzen wir in diesem Absatz voraus, daß I endlich ist, so ergeben sich einige einfache Folgerungen: Für endliches I stimmen sowohl die beiden Äquivalenzrelationen $\sim_0$ und $\sim$, als auch die Topologien $\mathcal{T}$, $\mathcal{T}_\infty$ und $\mathcal{T}_1$ überein. Als endliche Mengen sind die Äquivalenzklassen kompakt und abgeschlossen. Für eine Menge $A \subset E_I^b$ ist $\bigcup_{\sigma \in \mathcal{T}(I)} \sigma(A)$ die saturierte Hülle von A bzgl. $\sim$. Da $\sigma \in \mathcal{T}(I)$ jeweils eine topologische Abbildung von E_I^b auf E_I^b induziert, ist für offenes A die saturierte Hülle als Vereinigung offener Mengen offen und für abgeschlossenes A die saturierte Hülle als endliche Vereinigung abgeschlossener Mengen abgeschlossen. Versieht man dann den (mengentheoretischen) Quotientenraum $E_{I/\sim}^b$ mit der Quotiententopologie, so ist die natürliche Projektion $\pi: E_I^b \rightarrow E_{I/\sim}^b$ stetig, offen und abgeschlossen. Viele topologische Eigenschaften übertragen sich von E_I^b auf $E_{I/\sim}^b$, insbesondere die Metrisierbarkeit, vgl. Balachandran (1955).

Im Falle einer abzählbar unendlichen Indexmenge I kann man diese Schlußfolgerungen nicht ziehen. π erweist sich zwar als offen (Satz 2 unten), ist aber i.a. nicht abgeschlossen, vgl. Beispiel 3 unten; trotzdem ist $E_{I/\sim}^b$ als Quotientenraum von $(E_I^b, \mathcal{T}_\infty)$ bzw. $(E_I^b, \mathcal{T}_1)$ bzgl. $\sim$ metrisierbar (Satz 3 unten). Für den Nachweis dieser Sätze ist der folgende Hilfssatz wichtig.

Hilfssatz

Es sei $I = \mathbb{N}$, $(\mu_i)_I \in E_I^b$ und $(\nu_i)_I \in E_I^b$, $\mathcal{T}^* \in \{\mathcal{T}_1, \mathcal{T}_\infty\}$. Genau dann existiert eine injektive (bzw. bijektive) Abbildung $\sigma: I_\nu \rightarrow I_\mu$ mit $\mu_{\sigma i} = \nu_i$ für $i \in I_\nu$, wenn eine Folge von Permutationen $(\sigma_n)_{n \in \mathbb{N}}$ in $\mathcal{T}_0(I)$ mit $\sigma_n(\mu_i)_I \rightarrow (\nu_i)_I$ bzgl. $\mathcal{T}$ (bzw. bzgl. $\mathcal{T}^*$) existiert.

Beweis " $\Rightarrow$ "

a) σ sei injektiv:

$(U_n)_N$ sei eine abzählbare Umgebungsbasis von $0 \in \mathcal{M}$ bzgl. der schwachen Topologie $w(X)$ auf $\mathcal{M}$ mit $U_n \supset U_{n+1}$ für $n \in \mathbb{N}$.

Wegen $\mu_i \rightarrow 0$ für $i \rightarrow \infty$ existieren zu jedem $n \in \mathbb{N}$ abzählbar viele Indizes $j(n) \in I$ mit $\mu_{j(n)} \in U_n$, also für $i \in \{1, 2, \dots, n\} \setminus I_p$ jeweils verschiedene $j(n, i) \in I \setminus \sigma(\{1, 2, \dots, n\} \cap I_p)$ mit

$\mu_{j(n, i)} \in U_n$. Setze

$$\sigma_n^* i := \begin{cases} \sigma i & \text{für } i \in \{1, \dots, n\} \cap I_p \\ j(n, i) & \text{für } i \in \{1, \dots, n\} \setminus I_p. \end{cases}$$

Nach Konstruktion ist $\sigma_n^*: \{1, \dots, n\} \rightarrow \mathbb{N}$ eine injektive Abbildung, da σ eine injektive Abbildung ist.

Sei $K_n := \max_{i=1}^n \sigma_n^* i$; dann gibt es eine bijektive Fortsetzung $\sigma'_n \in \mathcal{S}(\{1, \dots, K_n\})$ von σ_n^* . Die Permutationen $\sigma_n \in \mathcal{S}_0(\mathbb{N})$ mit

$$\sigma_n^i := \begin{cases} \sigma_n^* i & \text{für } i \leq K_n \\ i & \text{für } i > K_n \end{cases}$$

erfüllen, wie im folgenden gezeigt wird, die Behauptung des

Satzes: $\sigma_n(\mu_i)_I \rightarrow (\mu_i)_I$ bzgl. $\mathcal{Z}$.

Für $i \in I$ und $i \leq n$ gilt $i \leq n \leq K_n$, also auch

$$\sigma_n i = \sigma'_n i = \sigma_n^* i.$$

1. Fall: $i \in I_p$; dann gilt $\sigma_n^* i = \sigma i$, also $\mu_{\sigma_n i} = \mu_{\sigma i} = \mu_i$.

Aus $\mu_{\sigma_n i} = \mu_i$ für $n \geq i$ folgt sofort $\mu_{\sigma_n i} \rightarrow \mu_i$ für $n \rightarrow \infty$.

2. Fall: $i \in I \setminus I_p$, also $\mu_i = 0$. Es folgt $\sigma_n^* i = j(n, i)$ mit

$$\mu_{j(n, i)} \in U_n, \text{ also } \mu_{\sigma_n i} \in U_n \text{ für } n \geq i.$$

Daher gibt es zu jedem $i \in I \setminus I_p$ und zu jedem U_m ein

$n_0 := \max(i, m)$ mit $\mu_{\sigma_n i} \in U_n \subset U_m$ für alle $n \geq n_0$, also folgt auch in diesem Falle $\mu_{\sigma_n i} \rightarrow 0 = \mu_i$ für $n \rightarrow \infty$.

Zusammen gilt $\sigma_n(\mu_i)_I \rightarrow (\mu_i)_I$ bzgl. $\mathcal{Z}$.

b) Ist σ sogar bijektiv, so gilt wegen Bemerkung 9 c)

$$\Sigma^X \sigma_n(\mu_i)_I = \Sigma^X(\mu_i)_I = \Sigma^X(\nu_i)_I.$$

Die Gesamtmasse bleibt unter σ_n also erhalten; mit Satz 1, a) gilt also auch

$$\sigma_n(\mu_i)_I \longrightarrow (\nu_i)_I \quad \text{bzgl. } \check{\tau}_\infty \text{ und } \check{\tau}_1.$$

"<="

a) Es gelte $\sigma_n(\mu_i)_I \longrightarrow (\nu_i)_I$ bzgl. $\check{\tau}$.

Da 0 der einzige Häufungspunkt von $\{\mu_i : i \in I\}$ ist und

$\mu_{\sigma_n i} \longrightarrow \nu_i$ für jedes $i \in I$ gilt, wird $(\mu_{\sigma_n i})_{n \in \mathbb{N}}$ schließlich für jedes $i \in I_\mu$ konstant. Wegen $\left| \{j \in I_\mu : \mu_j = \nu_i\} \right| < \infty$ können wir o.E. annehmen, daß $(\sigma_n i)_{n \in \mathbb{N}}$ schließlich konstant

wird für $i \in I_\mu$, nötigenfalls kann man ja sukzessive zu Teilfolgen von $(\sigma_n)_{n \in \mathbb{N}}$ und dann zu der Diagonalfolge übergehen.

Die Abbildung $\sigma^k : \{1, 2, \dots, k\} \cap I_\mu \longrightarrow I_\mu$ definiert durch

$\sigma^k i := \lim_n \sigma_n i$ - oder äquivalent durch $\sigma^k := \sigma_n / \{1, \dots, k\} \cap I_\mu$ für genügend großes n - ist mit σ_n für jedes $k \in \mathbb{N}$ injektiv.

Nach Konstruktion gilt für $k^* \geq k$ $\sigma^{k^*} / \{1, \dots, k\} \cap I_\mu = \sigma^k$.

Also ist $\sigma : I_\mu \longrightarrow I_\mu$ definiert durch $\sigma i := \sigma^k i$ für $k \geq i$, $i \in I_\mu$, eine injektive Abbildung. Weiter gilt $\mu_{\sigma i} = \lim_n \mu_{\sigma_n i} = \nu_i$ für $i \in I_\mu$.

b) Für den letzten Teil genügt es zu zeigen:

I) Aus $\sigma_n(\mu_i)_I \longrightarrow (\nu_i)_I$ bzgl. $\check{\tau}_\infty$, folgt auch

$$\sigma_n(\mu_i)_I \longrightarrow (\nu_i)_I \quad \text{bzgl. } \check{\tau}_1.$$

II) Aus $\sigma_n(\mu_i)_I \longrightarrow (\nu_i)_I$ bzgl. $\check{\tau}_1$ folgt die Surjektivität der in "<=" a) konstruierten Abbildung $\sigma : I_\mu \longrightarrow I_\mu$.

Zu I) Es gelte nun $\sigma_n(\mu_i)_I \longrightarrow (\nu_i)_I$ bzgl. $\check{\tau}_\infty$, d.h.

$\sup_I d(\mu_{\sigma_n i}, \nu_i) \longrightarrow 0$. Da $\sigma_n \in \mathcal{T}_0(\mathbb{N})$ bijektiv ist, ist dies gleichbedeutend mit $\sup_I d(\mu_i, \mu_{\sigma_n^{-1} i}) \longrightarrow 0$ oder

$\sigma_n^{-1}(\mu_i)_I \longrightarrow (\mu_i)_I$ bzgl. $\mathcal{T}_\infty$. Da Σ^X unter $\mathcal{D}_0(N)$ invariant ist, gilt unter Berücksichtigung von Bemerkung 4:

$$\Sigma^X(\mu_i)_I = \lim_n \Sigma^X \sigma_n(\mu_i)_I \geq \Sigma^X(\mu_i)_I = \lim_m \Sigma^X \sigma_m^{-1}(\mu_i)_I \geq \Sigma^X(\mu_i)_I;$$

die Gesamtmasse bleibt erhalten; mit Satz 1, a) folgt

$$\sigma_n(\mu_i)_I \longrightarrow (\mu_i)_I \text{ bzgl. } \mathcal{T}_1.$$

Zu II) Es gelte $\sigma_n(\mu_i)_I \longrightarrow (\mu_i)_I$ bzgl. $\mathcal{T}_1$. Wir nehmen für einen indirekten Beweis an, das oben konstruierte $\sigma: I_\mu \longrightarrow I_{\mu^*}$ sei nicht surjektiv. Setze

$$\mu_i^* := \begin{cases} \mu_i & \text{für } i \in \sigma(I_\mu) \\ 0 & \text{sonst,} \end{cases}$$

dann gilt für $I_{\mu^*} := \{i \in I : \mu_i^* \neq 0\}$ $\sigma: I_\mu \longrightarrow I_{\mu^*}$ ist bijektiv. Also existiert eine Folge $\sigma_n^* \in \mathcal{D}_0(N)$ mit

$$\sigma_n^*(\mu_i^*)_I \longrightarrow (\mu_i)_I \text{ bzgl. } \mathcal{T}_1. \quad \text{Also gilt:}$$

$$\Sigma^X(\mu_i)_I = \lim_n \Sigma^X \sigma_n^*(\mu_i^*)_I = \Sigma^X(\mu_i^*)_I < \Sigma^X(\mu_i)_I.$$

Dies aber steht im Widerspruch zu $\sigma_n(\mu_i)_I \longrightarrow (\mu_i)_I$ bzgl. $\mathcal{T}_1$ wegen Satz 1, a). Also ist $\sigma: I_\mu \longrightarrow I_{\mu^*}$ surjektiv.

Q.e.d.

Bemerkung

10. Für $I = \mathbb{N}$ folgt aus dem vorhergehenden Hilfssatz

$$(\pi_0(\mu_i)_I)^{a(\mathcal{T}_\infty)} = (\pi_0(\mu_i)_I)^{a(\mathcal{T}_1)} = \pi(\mu_i)_I.$$

Da diese Menge abgeschlossen ist bzgl. $\mathcal{T}_\infty$ und $\mathcal{T}_1$, folgt weiter

$$\pi(\mu_i)_I = (\pi(\mu_i)_I)^{a(\mathcal{T}_1)} = (\pi(\mu_i)_I)^{a(\mathcal{T}_\infty)}.$$

Für endliches I gilt dies natürlich auch, weil sich die Topologien und die Äquivalenzrelationen jeweils nicht unterscheiden. Aus dem Hilfssatz liest man weiter für $I = \mathbb{N}$ ab:

$$(\pi(\mu_i)_I)^{a(\tilde{\gamma})} = (\pi_o(\mu_i)_I)^{a(\tilde{\gamma})} =$$

$$= \{(\mu_i)_I \in E_I^b : \text{es existiert eine injektive Abbildung } \mathcal{J}: I_\mu \rightarrow I_\mu \\ \text{mit } \mu_i = \mu_{\mathcal{J}i} \quad \text{für } i \in I_\mu\}$$

Insbesondere gilt $(0)_I \in (\pi(\mu_i)_I)^{a(\tilde{\gamma})}$ für alle $(\mu_i)_I \in E_I^b$.

Es sei nun $\tilde{\gamma}^* \in \{\tilde{\gamma}, \tilde{\gamma}_\infty, \tilde{\gamma}_1\}$ und $\sim^* \in \{\sim, \sim\}$;

$\pi^*: E_I^b \longrightarrow E_{I/\sim^*}^b$ sei jeweils die natürliche Projektion,
also $\pi^* \in \{\pi_o, \pi\}$. Dann können wir definieren:

Definition

$$(E_I^b, \tilde{\gamma}^*) / \sim^*, (E_I^{\leq}, \tilde{\gamma}^* \cap E_I^{\leq}) / \sim^* \quad \text{und} \quad (E_I^1, \tilde{\gamma}^* \cap E_I^1) / \sim^*$$

seien die topologischen Quotientenräume.

Die folgenden beiden Sätze beschreiben die topologische Struktur dieser Räume.

Satz 3

a) Die natürliche Projektion

$$\pi^*: (E_I^b, \tilde{\gamma}^*) \longrightarrow (E_I^b, \tilde{\gamma}^*) / \sim^* \quad \text{ist offen.}$$

b) $(E_I^{\leq}, \tilde{\gamma}^* \cap E_I^{\leq}) / \sim^*$ und $(E_I^1, \tilde{\gamma}^* \cap E_I^1) / \sim^*$ tragen die

Spur von $(E_I^b, \tilde{\gamma}^*) / \sim^*$.

c) Eine Folge $(\pi^*(\mu_i^n)_I)_{n \in \mathbb{N}}$ in $(E_I^b, \tilde{\gamma}^*) / \sim^*$ konvergiert genau dann gegen $\pi^*(\mu_i)_I \in E_{I/\sim^*}^b$, wenn zu jedem Repräsentanten $(\mu_i)_I \in \pi^*(\mu_i)_I$ eine Folge von Repräsentanten $((\mu_i^n)_I)_{n \in \mathbb{N}}$ mit $(\mu_i^n)_I \in \pi^*(\mu_i^n)_I$ existiert, die gegen $(\mu_i)_I$ bzgl. $\tilde{\gamma}^*$ konvergiert.

d) Sei $I = \mathbb{N}$. Dann gilt:

i) $(E_I^b, \tilde{\gamma}_\infty) / \sim$ und $(E_I^b, \tilde{\gamma}_1) / \sim$ sind T_1 -Räume.

ii) $(E_I^b, \tilde{\gamma}) / \sim$ ist ein T_0 -Raum, aber kein T_1 -Raum,

alle stetigen reellen Funktionen auf $(E_I^b, \mathcal{F})/\sim$ sind konstant.

iii) $(E_I^b, \mathcal{F}^*)/\sim$ ist kein T_0 -Raum.

Beweis:

a) Mit G ist auch $G_0 := \bigcup_{\sigma \in \mathcal{F}_0} \sigma(G)$ $\mathcal{F}^*$ -offen in E_I^b ,

da $\sigma \in \mathcal{F}_0(I)$ topologische Abbildungen $\sigma: E_I^b \longrightarrow E_I^b$

bzgl. $\mathcal{F}^*$ liefern. Für endliches I ergibt sich hieraus schon die Behauptung. Sei nun $I = \mathbb{N}$. Es genügt zu zeigen: G_0 ist die saturierte Hülle von G unter $\sim$. Sei $(\mu_i)_I \in G$ und

$(\nu_i)_I \sim (\mu_i)_I$, also $(\nu_i)_I \sim (\mu_i)_I$; dann existiert eine Folge $(\sigma_n)_{n \in \mathbb{N}}$ in $\mathcal{F}_0(\mathbb{N})$ mit $\sigma_n(\nu_i)_I \longrightarrow (\mu_i)_I \in G$ bzgl. $\mathcal{F}_1$, also auch bzgl. $\mathcal{F}^*$. Für große $n \in \mathbb{N}$ gilt dann $\sigma_n(\nu_i)_I \in G$, da G $\mathcal{F}^*$ -offen ist, also $(\nu_i)_I \in \sigma_n^{-1}(G) \subset G_0$. G_0 enthält mit $(\mu_i)_I$ also die ganze Äquivalenzklasse $\pi^*(\mu_i)_I$. Da für $(\nu_i)_I \in G_0$ andererseits ein $\sigma \in \mathcal{F}_0(\mathbb{N})$ und ein $(\mu_i)_I \in G$ existiert mit $(\nu_i)_I = \sigma(\mu_i)_I$, liegt G_0 in der saturierten Hülle von G .

b) folgt aus Bourbaki (1966), Seite 53, Satz 4, da $E_I^{\leq}$ und E_I^1 in E_I^b bzgl. $\sim$ saturiert sind und π^* jeweils offen ist.

c) " $\Leftarrow$ " folgt aus der Stetigkeit von π^* .

" $\Rightarrow$ " folgt aus Lemma 3 von Balachandran (1955), da π^* eine offene, surjektive Abbildung ist und E_I^b bzgl. $\mathcal{F}^*$ das erste Abzählbarkeitsaxiom erfüllt.

d) i) Da wegen Bemerkung 10 die Menge $\pi(\mu_i)_I$ in $(E_I^b, \mathcal{F}_\infty)$ und $(E_I^b, \mathcal{F}_1)$ abgeschlossen ist für jedes $(\mu_i)_I \in E_I^b$, sind die Räume $(E_I^b, \mathcal{F})/\sim$ und $(E_I^b, \mathcal{F}_1)/\sim$ jeweils T_1 -Räume.

ii) Andererseits ergibt sich aus Bemerkung 10 sofort:

$(0)_I \in (\pi(\mu_i)_I)^{a(\mathcal{F})}$ für alle $(\mu_i)_I \in E_I^b$, d.h. $\pi(0)_I$ liegt

in jeder abgeschlossenen Menge A von $(E_I^b, \mathcal{F})/\sim$ mit $A \neq \emptyset$. Also ist $(E_I^b, \mathcal{F})/\sim$ kein T_1 -Raum; weiterhin ist jede stetige, reelle Funktion auf $(E_I^b, \mathcal{F})/\sim$ konstant. Aus Bemerkung 10 ergibt sich aber auch:

$(\pi(\mu_i)_I)^a(\mathcal{F}) = (\pi(\nu_i)_I)^a(\mathcal{F})$ gilt genau dann, wenn $(\mu_i)_I \sim (\nu_i)_I$ gilt. Folglich ist $(E_I^b, \mathcal{F})/\sim$ ein T_0 -Raum.

iii) Wegen $(\pi_o(\mu_i)_I)^a(\mathcal{F}^*) \supset \pi(\mu_i)_I$ gilt für $(\mu_i)_I \sim (\nu_i)_I$ in E_I^b : Der Punkt $\pi_o(\mu_i)_I$ liegt in genau den gleichen abgeschlossenen Mengen $A \subset (E_I^b, \mathcal{F}^*)/\sim$ wie $\pi_o(\nu_i)_I$.

Topologisch kann man $\pi_o(\mu_i)_I$ und $\pi_o(\nu_i)_I$ also nicht unterscheiden; $(E_I^b, \mathcal{F}^*)/\sim$ ist also kein T_0 -Raum. Der zugehörige T_0 -Raum - Übergang zu den Äquivalenzklassen topologisch nicht unterscheidbarer Elemente - ist offenbar bis auf Isomorphie der Raum $(E_I^b, \mathcal{F}^*)/\sim$.

Q.e.d.

Im nächsten Satz wird die Metrisierbarkeit der Quotientenräume untersucht. Dazu benötigen wir einige Bezeichnungen:

Bezeichnungen

Zu einer Metrik D^* auf E_I^b sei die Hausdorffmetrik auf $\mathcal{P}_0 := \mathcal{P} \setminus \{\emptyset\}$

$D_H^*: \mathcal{P}_0(E_I^b) \times \mathcal{P}_0(E_I^b) \longrightarrow \bar{\mathbb{R}}$ mit $D_H^*(A, B) :=$

$$\max \left(\sup_{(\mu_i)_I \in A} \inf_{(\nu_i)_I \in B} D^*((\mu_i)_I, (\nu_i)_I), \sup_{(\nu_i)_I \in B} \inf_{(\mu_i)_I \in A} D^*((\mu_i)_I, (\nu_i)_I) \right)$$

(eine "Fasthalbmetrik") eingeführt. Analog bezeichne wie

üblich $D^*: \mathcal{P}_0(E_I^b) \times \mathcal{P}_0(E_I^b) \longrightarrow \mathbb{R}$ mit

$$D^*(A, B) := \inf_{(\mu_i)_I \in A} \inf_{(\nu_i)_I \in B} D^*((\mu_i)_I, (\nu_i)_I)$$

auch den Abstand zwischen Mengen in $\mathcal{P}_0(E_I^b)$.

Satz 4

a) Für $(\mu_i)_I, (\nu_i)_I \in E_I^b$ und $D^* \in \{D_\infty, D_1\}$ gilt:

$$D_H^*(\pi(\mu_i)_I, \pi(\nu_i)_I) = D_H^*(\pi_o(\mu_i)_I, \pi_o(\nu_i)_I) =$$

$$D^*(\pi(\mu_i)_I, \pi(\nu_i)_I) = D^*(\pi_o(\mu_i)_I, \pi_o(\nu_i)_I) =$$

$$D^*(\pi(\mu_i)_I, \{(\nu_i)_I\}) = D^*(\pi_o(\mu_i)_I, \{(\nu_i)_I\}) .$$

b) Die Hausdorffmetrik (eine "Fasthalbmetrik") zu D_∞

bzw. D_1 induziert auf $E_I^b / \sim$ die Topologie von $(E_I^b, \mathcal{T}_\infty) / \sim$

bzw. $(E_I^b, \mathcal{T}_1) / \sim$. Insbesondere sind die Räume $(E_I^b, \mathcal{T}_\infty) / \sim$ und

$(E_I^b, \mathcal{T}_1) / \sim$ durch die Hausdorffmetrik metrisierbar.

Beweis:

a) Für eine Metrik D^* , die unter $\mathcal{T}_o(I)$ invariant ist, gilt:

$$D_H^*(\pi_o(\mu_i)_I, \pi_o(\nu_i)_I) =$$

$$= \max \left(\sup_{\sigma \in \mathcal{T}_o(I)} \inf_{\tau \in \mathcal{T}_o(I)} D^*(\sigma(\mu_i)_I, \tau(\nu_i)_I), \sup_{\tau \in \mathcal{T}_o(I)} \inf_{\sigma \in \mathcal{T}_o(I)} D^*(\sigma(\mu_i)_I, \tau(\nu_i)_I) \right)$$

$$= \max \left(\sup_{\sigma} \inf_{\tau} D^*((\mu_i)_I, \sigma^{-1}\tau(\nu_i)_I), \sup_{\tau} \inf_{\sigma} D^*(\tau^{-1}\sigma(\mu_i)_I, (\nu_i)_I) \right)$$

$$= \max \left(\inf_{\tau} D^*((\mu_i)_I, \tau(\nu_i)_I), \inf_{\sigma} D^*(\sigma(\mu_i)_I, (\nu_i)_I) \right) =$$

$$= \inf_{\tau} D^*((\mu_i)_I, \tau(\nu_i)_I) = \inf_{\sigma} D^*(\sigma(\mu_i)_I, (\nu_i)_I) =$$

$$D^*(\pi_o(\mu_i)_I, \{(\nu_i)_I\}) = \inf_{\sigma} \inf_{\tau} D^*(\sigma(\mu_i)_I, \tau(\nu_i)_I) =$$

$$= D^*(\pi_o(\mu_i)_I, \pi_o(\nu_i)_I) .$$

Wegen $D^*(A, B) = D^*(A^a, B^a)$ und $D_H^*(A, B) = D_H^*(A^a, B^a)$ bzgl.

der von D^* induzierten Topologien $\mathcal{T}_\infty$ bzw. $\mathcal{T}_1$ folgt mit

Bemerkung 10 die Aussage a).

b) Aus a) folgt für $\pi^* \in \{\pi_o, \pi\}$:

Eine Folge $(\pi^*(\mu_i^n)_I)_{n \in \mathbb{N}}$ in $E_I^b / \sim$ konvergiert genau dann bzgl.

D_H^* gegen $\pi^*(\mu_i)_I \in E_I^b / \sim$, wenn zu jedem Repräsentanten $(\mu_i)_I$ von $\pi^*(\mu_i)_I$ eine Folge von Repräsentanten $((\mu_i^n)_I)_{\mathbb{N}}$ mit $(\mu_i^n)_I \in \pi^*(\mu_i^n)_I$ existiert, die gegen $(\mu_i)_I$ bzgl. D^* konvergiert. Da $(E_I^b, \mathcal{V}^*) / \sim$ das erste Abzählbarkeitsaxiom erfüllt - π^* ist eine offene, stetige und surjektive Abbildung, vgl. Balachandran (1955), Lemma 1 - folgt der erste Teil von Aussage b) aus Satz 3, c). Da $D_H^* = D^*$ auf $E_I^b / \sim$ gilt, also insbesondere endlich ist, und weil $(E_I^b, \mathcal{V}_\infty) / \sim$ und $(E_I^b, \mathcal{V}_1) / \sim$ wegen Satz 3, d) T_1 -Räume sind, folgt auch die zweite Aussage von b).

Q.e.d.

Bemerkungen

11. Auf dem Raum $E_I^b / \sim$ ist D_H^* nur eine Halbmetrik; der metrische Raum, der durch Faktorisierung nach der Äquivalenzrelation $D_H^*(\pi_0(\mu_i)_I, \pi_0(\nu_i)_I) = 0$ entsteht, ist offensichtlich zu $(E_I^b / \sim, D_H^*)$ isometrisch. Vgl. hierzu auch den letzten Satz im Beweis von Satz 3.

12. Die Äquivalenzrelation $\approx$ auf E_I^b , die durch $(\mu_i)_I \approx (\nu_i)_I : \Leftrightarrow$ Es existiert $\mathfrak{J} \in \mathcal{J}(I)$ mit $\mathfrak{J}(\mu_i)_I = (\nu_i)_I$ definiert wird, liegt zwischen den beiden Äquivalenzrelationen $\sim_0$ und $\sim$, d.h. es gilt: $\sim_0 \subset \approx \subset \sim$. Für endliches I stimmt sie also mit den beiden überein. Zwei Elemente $(\mu_i)_I, (\nu_i)_I \in E_I^b$ mit $(\mu_i)_I \not\approx (\nu_i)_I$ aber $(\mu_i)_I \sim (\nu_i)_I$ unterscheiden sich durch die "Anzahl der Nullen", d.h. $|I \setminus I_\mu| \neq |I \setminus I_\nu|$.

Der Quotientenraum $(E_I^b, \mathcal{V}^*) / \sim$ wird durch die Hausdorffmetrik analog zu $(E_I^b, \mathcal{V}^*) / \sim_0$ halbmetrisiert, er ist kein T_0 -Raum.

Im folgenden beschränken wir uns daher auf die Quotientenräume bzgl. der Äquivalenzrelation $\sim$.

Bezeichnungen

Die Quotiententopologie der Räume $(E_I^b, \tilde{\tau})/\sim$, $(E_I^b, \tilde{\tau}_\infty)/\sim$ und $(E_I^b, \tilde{\tau}_1)/\sim$ bezeichnen wir mit $\tilde{\tau}$, $\tilde{\tau}_\infty$ bzw. $\tilde{\tau}_1$.

Σ^X und Σ sind auf E_I^b invariant unter $\sim$ definiert; daher existieren Abbildungen $\tilde{\Sigma}^X: E_I^b/\sim \rightarrow \mathbb{R}$ und $\tilde{\Sigma}: E_I^b/\sim \rightarrow \mathcal{M}$ mit $\tilde{\Sigma}^X \circ \pi = \Sigma^X$ und $\tilde{\Sigma} \circ \pi = \Sigma$. Da keine Mißverständnisse zu befürchten sind, schreiben wir weiter Σ^X statt $\tilde{\Sigma}^X$ und Σ statt $\tilde{\Sigma}$.

Die Quotientenmengen $\tilde{E}_I^\leq := E_I^\leq/\sim$ bzw. $\tilde{E}_I^1 := E_I^1/\sim$ heißen der Raum aller Clusterungen auf X mit der Indexmenge I , bzw. der Raum aller Clusterungen ohne Masseverlust auf X mit der Indexmenge I . Die Einschränkungen der Topologien $\tilde{\tau}$, $\tilde{\tau}_\infty$ und $\tilde{\tau}_1$ auf $\tilde{E}_I^\leq$ und $\tilde{E}_I^1$ bezeichnen wir wieder mit $\tilde{\tau}$, $\tilde{\tau}_\infty$ und $\tilde{\tau}_1$, wenn keine Mißverständnisse zu befürchten sind.

Ist die natürliche Projektion von einem metrisierbaren, topologischen Raum auf einen Quotientenraum offen und abgeschlossen, so folgt die Metrisierbarkeit des Quotientenraumes aus bekannten Sätzen (Balachandran (1955)). Das folgende Beispiel zeigt, daß die natürliche Projektion π im vorliegenden Fall i.a. nicht abgeschlossen ist.

Beispiel

2. Es sei $I = \mathbb{N}$, $X = \mathbb{R}$, $\mu_i^0 = 0$ für $i \in I$.

$$\mu_i^n := \begin{cases} \delta_i & \text{für } i = n \\ 0 & \text{sonst} \end{cases}$$

$$F := \{(\mu_i^n)_I : n \in \mathbb{N} \cup \{0\}\}.$$

Wegen $D((\mu_i^n)_I, (\mu_i^0)_I) \rightarrow 0$ für $n \rightarrow \infty$ ist F bzgl. $\tilde{\tau}$

und damit erst recht bzgl. $\tilde{\tau}_\infty$ und $\tilde{\tau}_1$ abgeschlossen in E_I^b .

$$\{(\delta_{1/n}, 0, 0, \dots) : n \in \mathbb{N}\} \subset \pi^{-1} \pi(F).$$

Diese Folge konvergiert bzgl. D_1 gegen $(\delta_0, 0, 0, \dots)$.

Aber es gilt auch: $(\delta_0, 0, 0, \dots) \notin \pi^{-1} \pi(F)$.

$\pi(F)$ ist also bzgl. keiner der Topologien $\tilde{\tau}$, $\tilde{\tau}_\infty$, $\tilde{\tau}_1$ abgeschlossen.

Bemerkungen

13. Aus der universellen Eigenschaft von Quotientenräumen ergeben sich die beiden folgenden Aussagen:

$$\tilde{\tau} \subset \tilde{\tau}_\infty \subset \tilde{\tau}_1.$$

Die Abbildung $\Sigma^X: E_I^b \longrightarrow \mathbb{R}$ ist bzgl. $\tilde{\tau}_1$ stetig, bzgl. $\tilde{\tau}$ und $\tilde{\tau}_\infty$ halbstetig nach unten, also ist auch $\Sigma^X: E_{I/\sim}^b \longrightarrow \mathbb{R}$ bzgl. $\tilde{\tau}_1$ stetig, bzgl. $\tilde{\tau}$ und $\tilde{\tau}_\infty$ halbstetig nach unten.

Die Abbildung $\Sigma: E_I^b \longrightarrow \mathcal{M}$ ist bzgl. $\tilde{\tau}_1$ stetig, bzgl. $\tilde{\tau}$ und $\tilde{\tau}_\infty$ nicht stetig, vgl. Beispiele 1 und 2. Also ist auch $\Sigma: E_{I/\sim}^b \longrightarrow \mathcal{M}$ bzgl. $\tilde{\tau}_1$ stetig, bzgl. $\tilde{\tau}$ und $\tilde{\tau}_\infty$ nicht stetig.

14. Die Topologien $\tilde{\tau}$, $\tilde{\tau}_\infty$ und $\tilde{\tau}_1$ stimmen auf E_I^1 überein (Satz 1, b)). Es gilt daher auch: $(E_I^1, \tilde{\tau} \cap E_I^1)_{/\sim} = (E_I^1, \tilde{\tau}_\infty \cap E_I^1)_{/\sim} = (E_I^1, \tilde{\tau}_1 \cap E_I^1)_{/\sim}$ und wegen Satz 3, b) ist dies die Spur der entsprechenden Quotiententopologie auf $E_{I/\sim}^b$. Die Topologien stimmen also auf $\tilde{E}_I^1$ überein.

15. Mit Satz 3 und 4 läßt sich Satz 1 auf die Quotientenräume übertragen:

a) Eine Folge $((\tilde{\mu}_i^n)_{i \in \mathbb{N}})_{n \in \mathbb{N}}$ in $\tilde{E}_I^b$ konvergiert genau dann bzgl. $\tilde{\tau}_1$ gegen $(\tilde{\mu}_i)_{i \in \mathbb{N}} \in \tilde{E}_I^b$, wenn $((\tilde{\mu}_i^n)_{i \in \mathbb{N}})_{n \in \mathbb{N}}$ bzgl. $\tilde{\tau}$ gegen $(\tilde{\mu}_i)_{i \in \mathbb{N}}$ konvergiert und zusätzlich die Gesamtmasse konvergiert:

$$\Sigma^X(\tilde{\mu}_i^n)_{i \in \mathbb{N}} \longrightarrow \Sigma^X(\tilde{\mu}_i)_{i \in \mathbb{N}}.$$

b) In $\tilde{E}_I^\leq$ stimmen die Umgebungssysteme der Punkte aus $\tilde{E}_I^1$ bzgl. der drei Topologien $\tilde{\tau}$, $\tilde{\tau}_\infty$ und $\tilde{\tau}_1$ überein.

Für den folgenden Satz benötigen wir folgende Definition:

Definition

Für $\mathcal{K} = \mathcal{M}^1$ sei

$$\begin{aligned}\tilde{E}(\mathcal{K}) &:= \tilde{E}_I^\leq(\mathcal{K}) := \left\{ (\tilde{\mu}_i)_I \in \tilde{E}_I^\leq : \Sigma(\tilde{\mu}_i)_I \in \mathcal{K} \leq \right\} \\ &= \left\{ (\tilde{\mu}_i)_I \in \tilde{E}_I^\leq : \Sigma(\tilde{\mu}_i)_I \leq \mu \quad \text{für ein } \mu \in \mathcal{K} \right\}.\end{aligned}$$

$\tilde{E}(\mathcal{K})$ ist die Menge aller Clusterungen von W-Maßen aus $\mathcal{K}$.

$$\tilde{E}(\mu) := \tilde{E}_I^\leq(\mu) := \tilde{E}_I^\leq(\{\mu\}).$$

Satz 5

Ist $\mathcal{K}$ kompakt, so ist auch $\tilde{E}(\mathcal{K})$ kompakt in $(\tilde{E}_I^\leq, \tilde{\tau}_\infty)$.

Beweis:

Da $(\tilde{E}_I^\leq, \tilde{\tau}_\infty)$ metrisierbar ist, genügt es, die Folgenkompaktheit zu zeigen. $((\tilde{\mu}_i^n)_I)_{n \in \mathbb{N}}$ sei eine Folge in $\tilde{E}(\mathcal{K})$.

Weil für $n \in \mathbb{N}$ $(\Sigma_I \tilde{\mu}_i^n)(X) \leq 1$, existieren nach Masse geordnete Repräsentanten $(\mu_i^n)_I$ von $(\tilde{\mu}_i^n)_I$, d.h. für $n \in \mathbb{N}$ und $i \in I$ gilt $\mu_i^n(X) \geq \mu_{i+1}^n(X)$. Nach Satz 2 existiert eine Teilfolge $((\mu_i^1)_I)_{i \in \mathbb{N}}$, die bzgl. $\tilde{\tau}$ gegen ein $(\mu_i)_I \in E(\mathcal{K})$

konvergiert. Es wird nun gezeigt, daß diese Teilfolge auch

bzgl. $\tilde{\tau}_\infty$ konvergiert. Zu beliebigem $\varepsilon > 0$ existiert ein $i_0 \in I$ mit $\mu_{i_0}^1(X) < \varepsilon/2$. Wegen $\mu_{i_0}^1(X) \longrightarrow \mu_{i_0}(X)$ gibt es ein $l_0 \in \mathbb{N}$, sodaß für alle $l \geq l_0$ gilt: $\mu_{i_0}^1(X) < \varepsilon$. Da $(\mu_i^1)_I$ nach Masse geordnet ist, gilt für dies $i_0 \in I$ und $l_0 \in \mathbb{N}$,

für jedes $i \geq i_0$ und jedes $l \geq l_0$ auch: $\mu_i^1(X) \leq \mu_{i_0}^1(X) < \varepsilon$.

Mit Bemerkung 6 c) folgt nun: $(\mu_i^1)_I \longrightarrow (\mu_i)_I$ bzgl. $\tilde{\tau}_\infty$.

Hieraus ergibt sich nun sofort die Behauptung des Satzes.

Das folgende Beispiel zeigt, daß dieser Satz nicht auch für $(\tilde{E}_I, \tilde{\tau}_1)$ gilt.

Beispiel

2*. Siehe Beispiel 2.

Da die Folge $((\tilde{\mu}_1^n)_I)_{n \in \mathbb{N}}$ bzgl. $\tilde{\tau}_\infty$ gegen $(\tilde{0})_I$ konvergiert, wäre $(\tilde{0})_I$ der Grenzwert jeder bzgl. $\tilde{\tau}_1$ konvergenten Teilfolge $((\tilde{\mu}_1^1)_I)_{1 \in \mathbb{N}}$. Es gilt aber

$$\lim_1 \Sigma^X(\tilde{\mu}_1^1)_I = \mu(X) \neq \Sigma^X(\tilde{0})_I.$$

Also gibt es keine konvergente Teilfolge: $\tilde{E}(\mu)$ ist nicht kompakt, aber $\{\mu\}$.

Definition

Sei $(\tilde{\mu}_1)_I \in \tilde{E}_I^{\leq}$ mit $\Sigma(\tilde{\mu}_1)_I \leq \mu \in \mathcal{M}^1$ und $(\mu_i)_I$ ein Repräsentant von $(\tilde{\mu}_1)_I$, so heißen die Teilmaße μ_i von μ

Cluster (der Clusterung $(\tilde{\mu}_1)_I$) von μ .

Sei $\mu = \frac{1}{N} \sum_{i=1}^N \delta_{x_i}$ ein empirisches Maß. $(\tilde{\mu}_1)_I \in \tilde{E}_I^{\leq}(\mu)$.

Dann heißt $(\tilde{\mu}_1)_I$ mengenartige Clusterung von μ , wenn für jedes Cluster μ_i der Clusterung $(\tilde{\mu}_1)_I$ von μ gilt:

$\mu_i = \frac{1}{N} \sum_{\text{gewisse } i} \delta_{x_i}$. Der Stichprobenumfang N ist dabei fest vorausgesetzt. Alle Cluster sind also zum gleichen Stichprobenumfang mengenartige Teilmaße von μ .

Satz 6

$(\mu^n)_{n \in \mathbb{N}}$ sei eine gegen μ konvergente Folge in $\mathcal{M}^1$.

Ferner sei $(\tilde{\mu}_1)_I \in \tilde{E}_I^{\leq}$.

a) Wenn es eine Folge $((\tilde{\mu}_1^n)_I)_{n \in \mathbb{N}}$ in $\tilde{E}_I^{\leq}$ gibt, die bzgl. $\tilde{\tau}_\infty$ gegen $(\tilde{\mu}_1)_I$ konvergiert und für die $\Sigma(\tilde{\mu}_1^n)_I \leq \mu^n$ für alle $n \in \mathbb{N}$ erfüllt ist, so ist $\Sigma(\tilde{\mu}_1)_I \leq \mu$.

- b) Ist andererseits $\Sigma(\tilde{\mu}_i)_I \leq \mu$, so existiert eine solche Folge $((\tilde{\mu}_i^n)_I)_{n \in \mathbb{N}}$, die sogar bzgl. $\tilde{\tau}_1$ konvergiert.
- c) Sind die Maße μ^n empirische Maße,

$$\mu^n = \frac{1}{N_n} \sum_{i=1}^{N_n} \delta_{x_{in}}, \quad n \in \mathbb{N}, \quad \text{mit } N_n \longrightarrow \infty, \quad \text{so}$$

kann man in b) die Clusterungen $((\tilde{\mu}_i^n)_I)_{n \in \mathbb{N}}$ mengenartig wählen.

Beweis:

a) In $\tilde{E}_I^{\leq}$ sei eine bzgl. $\tilde{\tau}_\infty$ gegen $(\tilde{\mu}_i)_I \in \tilde{E}_I^{\leq}$ konvergente Folge $((\tilde{\mu}_i^n)_I)_{n \in \mathbb{N}}$ mit $\Sigma(\tilde{\mu}_i^n)_I \leq \mu^n$ für alle $n \in \mathbb{N}$ gegeben. Dann existieren nach Satz 3, c) Repräsentanten $(\mu_i^n)_I$, $(\mu_i)_I$ mit $(\mu_i^n)_I \longrightarrow (\mu_i)_I$ bzgl. $\tilde{\tau}_\infty$ für $n \longrightarrow \infty$.

Für $f \in C(X)$, $f \geq 0$ gilt:

$$\begin{aligned} \int f d\mu &= \lim_n \int f d\mu^n \geq \liminf_n \int f d\Sigma(\mu_i^n)_I \geq \Sigma_{i \in I} \liminf_n \int f d\mu_i^n \geq \\ &\geq \Sigma_{i \in I} \int f d\mu_i; \end{aligned}$$

also $\Sigma_I \mu_i \leq \mu$ und $\Sigma(\tilde{\mu}_i)_I \leq \mu$.

b) und c) Sei $\nu := \Sigma(\tilde{\mu}_i)_I \leq \mu$. O.E. sei dabei $I = \mathbb{N}$. Nach I, Satz 2 existiert eine gegen ν konvergente Teilfolge $(\nu^n)_{n \in \mathbb{N}}$ von (mengenartigen) Teilmaßen mit $\nu^n \leq \mu^n$. $(\mu_i)_I$ sei ein Repräsentant von $(\tilde{\mu}_i)_I$. Zu μ_1 existiert nach I, Satz 2 eine gegen μ_1 konvergente Folge $(\mu_1^n)_{n \in \mathbb{N}}$ von (mengenartigen) Teilmaßen $\mu_1^n \leq \nu^n$. Nach Konstruktion gilt $\mu_1^n \longrightarrow \mu_1$, $\mu_1^n \leq \nu^n$, $\nu^n \longrightarrow \nu = \Sigma_I \mu_i$, also auch $\nu^n - \mu_1^n \longrightarrow \nu - \mu_1 = \Sigma_{i>1} \mu_i$, (und $\nu^n - \mu_1^n$ ist mengenartig).

Mittels vollständiger Induktion kann man nun wieder mit I, Satz 2 für jedes $k \in \mathbb{N}$ die Existenz einer Folge

$\mu_k^n \leq \nu^n - \Sigma_{i<k} \mu_i^n$ von (mengenartigen) Teilmaßen $\mu_k^n \leq \nu^n$ nachweisen mit $\mu_k^n \longrightarrow \mu_k$ für $n \longrightarrow \infty$ und $\Sigma_{i \leq k} \mu_i^n \leq \nu^n$.

Für die (mengenartigen) Folgen $((\mu_i^n)_I)_{n \in \mathbb{N}} \in E_I^{\leq}$ gilt nun
 $(\mu_i^n)_I \longrightarrow (\mu_i)_I$ bzgl. $\mathcal{T}$ und $\sum_I \mu_i^n \leq \nu^n$ für $n \in \mathbb{N}$.

Aus $\limsup_n \sum_I \mu_i^n(X) \leq \lim_n \nu^n(X) = \nu(X) = \sum_I \mu_i(X)$ und
der Halbstetigkeit nach unten von Σ^X bzgl. $\mathcal{T}$ (Bemerkung 4)
folgt $\sum_I \mu_i^n(X) \longrightarrow \sum_I \mu_i(X)$, die Gesamtmasse bleibt
erhalten. Satz 1 liefert nun die Konvergenz bzgl. $\mathcal{T}_1$,
also auch $(\tilde{\mu}_i^n)_I \longrightarrow (\tilde{\mu}_i)_I$ bzgl. $\tilde{\mathcal{T}}_1$.

Q.e.d.

Bei den folgenden Überlegungen sei J eine beliebige nicht
leere Teilmenge von I . Wir wollen die Räume $E_J^{\leq}$ in $E_I^{\leq}$ ein-
betten. Sei dazu $\text{inj}: E_J^{\leq} \longrightarrow E_I^{\leq}$ definiert durch
 $\text{inj}(\mu_j)_J := (\mu_i)_I$ mit $\mu_i = \mu_j$ für $i \in J$, $\mu_i = 0$
sonst. Die Abbildung inj ist injektiv, jeweils bzgl. der
Topologien $\mathcal{T}$, $\mathcal{T}_\infty$ und $\mathcal{T}_1$ auf $E_J^{\leq}$ bzw. auf $E_I^{\leq}$ stetig, und
bzgl. der Metriken D_∞ und D_1 auf $E_J^{\leq}$ und $E_I^{\leq}$ eine Isometrie.

$$\begin{array}{ccc} E_J^{\leq} & \xrightarrow{\text{inj}} & E_I^{\leq} \\ \pi_J \downarrow & & \downarrow \pi = \pi_I \\ \tilde{E}_J^{\leq} & \xrightarrow{\tilde{\text{inj}}} & \tilde{E}_I^{\leq} \end{array}$$

Es ist $(\mu_j)_J \sim (\nu_j)_J$ genau dann, wenn $\text{inj}(\mu_j)_J \sim \text{inj}(\nu_j)_J$,
oder $\pi \text{inj}(\mu_j)_J = \pi \text{inj}(\nu_j)_J$. Es gibt daher genau eine injek-
tive Abbildung $\tilde{\text{inj}}: \tilde{E}_J^{\leq} \longrightarrow \tilde{E}_I^{\leq}$, derart daß $\pi_I \circ \text{inj} = \text{inj} \circ \pi_J$
gilt. $\tilde{\text{inj}}$ ist stetig, weil $\tilde{E}_J^{\leq}$ die Finaltopologie unter π_J
trägt und $\pi_I \circ \text{inj}$ stetig ist.

Wir können nun definieren:

$$E_{J \subset I}^{\leq} := \text{inj}(E_J^{\leq}) \quad \text{und} \quad \tilde{E}_{J \subset I}^{\leq} := \tilde{\text{inj}}(\tilde{E}_J^{\leq}).$$

Kanonisch zerlegen wir inj in $\text{inj}_2 \circ \text{inj}_1$ und $\tilde{\text{inj}}$ in
 $\tilde{\text{inj}}_2 \circ \tilde{\text{inj}}_1$.

$$\begin{array}{ccccc}
 E_J^{\leq} & \xrightarrow{\text{inj}_1} & E_{J \subseteq I}^{\leq} & \xrightarrow{\text{inj}_2} & E_I^{\leq} \\
 \pi_J \downarrow & & \downarrow \pi_{J \subseteq I} & & \downarrow \pi = \pi_I \\
 \widetilde{E}_J^{\leq} & \xrightarrow{\widetilde{\text{inj}}_1} & \widetilde{E}_{J \subseteq I}^{\leq} & \xrightarrow{\widetilde{\text{inj}}_2} & \widetilde{E}_I^{\leq}
 \end{array}$$

Sei $\pi_{J \subseteq I}^0: E_{J \subseteq I}^{\leq} \longrightarrow \widetilde{E}_I^{\leq}$ die Einschränkung von π_I auf $E_{J \subseteq I}^{\leq}$

Wegen $\pi_{J \subseteq I}^0(E_{J \subseteq I}^{\leq}) = \pi_{J \subseteq I}^0 \text{inj}(E_J^{\leq}) = \widetilde{\text{inj}} \circ \pi_J(E_J^{\leq}) = \widetilde{\text{inj}}(\widetilde{E}_J^{\leq}) = \widetilde{E}_{J \subseteq I}^{\leq}$

existiert eine surjektive Abbildung $\pi_{J \subseteq I}: E_{J \subseteq I}^{\leq} \longrightarrow \widetilde{E}_{J \subseteq I}^{\leq}$

mit $\widetilde{\text{inj}}_2 \circ \pi_{J \subseteq I} = \pi_{J \subseteq I}^0$ - nur der Nachbereich der Funktion ist verschieden.

inj_1 und $\widetilde{\text{inj}}_1$ sind bijektiv, inj_2 und $\widetilde{\text{inj}}_2$ Einbettungen von Teilmengen. $\pi_{J \subseteq I} \circ \text{inj}_1 = \widetilde{\text{inj}}_1 \circ \pi_J$ gilt wegen $\pi_I \circ \text{inj} = \widetilde{\text{inj}} \circ \pi_J$.

Das gesamte Diagramm ist daher kommutativ.

$E_{J \subseteq I}^{\leq}$ und $\widetilde{E}_{J \subseteq I}^{\leq}$ trage die Initialtopologie unter inj_2

bzw. $\widetilde{\text{inj}}_2$. Dann sind sämtliche Abbildungen stetig, da inj

und $\widetilde{\text{inj}}$ stetig sind, $\pi_{J \subseteq I}$ ist stetig, da $\pi_I \circ \text{inj}_2$ stetig

ist und $\widetilde{E}_{J \subseteq I}^{\leq}$ die Initialtopologie unter $\widetilde{\text{inj}}_2$ (Spartopologie)

trägt. Wie man direkt sieht, ist inj_1 sogar eine Isometrie bezüglich D_∞ und D_1 , also ist inj_1 erst recht topologisch bzgl. $\mathcal{T}_\infty$ und $\mathcal{T}_1$. Direkt erkennt man auch, daß inj_1

bzgl. $\mathcal{T}$, der Produkttopologie, topologisch ist.

Wir können nun den folgenden Satz beweisen:

Satz 7

Die Abbildungen $\text{inj}: (E_J^{\leq}, \mathcal{T}^*) \longrightarrow (E_I^{\leq}, \mathcal{T}^*)$ und

$\widetilde{\text{inj}}: (\widetilde{E}_J^{\leq}, \widetilde{\mathcal{T}}^*) \longrightarrow (\widetilde{E}_I^{\leq}, \widetilde{\mathcal{T}}^*)$ sind für $\mathcal{T}^* \in \{\mathcal{T}, \mathcal{T}_\infty, \mathcal{T}_1\}$ und

$\widetilde{\mathcal{T}}^* \in \{\widetilde{\mathcal{T}}_\infty, \widetilde{\mathcal{T}}_1\}$ abgeschlossene Einbettungen.

Gilt $|J| = |I| = \infty$, so ist $\widetilde{\text{inj}}$ eine topologische Abbildung.

Beweis:

1. $E_{J \subset I}^{\leq}$ ist in $E_I^{\leq}$ abgeschlossen bzgl. $\mathcal{T}$ und erst recht bzgl. $\mathcal{T}_{\infty}$ und $\mathcal{T}_1$, da

$$E_{J \subset I}^{\leq} = \bigcap_{i \in I \setminus J} \text{pr}_i^{-1}(\{0\}).$$

$\tilde{E}_{J \subset I}^{\leq}$ ist in $\tilde{E}_I^{\leq}$ abgeschlossen bzgl. $\tilde{\mathcal{T}}$ und erst recht bzgl. $\tilde{\mathcal{T}}_{\infty}$ und $\tilde{\mathcal{T}}_1$. Hierzu genügt es zu zeigen:

$\pi^{-1}(\tilde{E}_{J \subset I}^{\leq}) = \{(\mu_i)_I \in E_I^{\leq} : |I_{\mu}| \leq |J|\}$ ist in $E_I^{\leq}$ abgeschlossen. Sei $(\mu_i)_I$ ein Berührungspunkt von $\pi^{-1}(\tilde{E}_{J \subset I}^{\leq})$,

also Limespunkt einer Folge von Elementen $(\mu_i^n)_I$ aus

$\pi^{-1}(\tilde{E}_{J \subset I}^{\leq})$. Für jedes endliche $J^* \subset I_{\mu}$ ist

$$U := \left(\prod_{J^*} (\mathcal{O} \leq \setminus \{0\}) \times \prod_{I \setminus J^*} \mathcal{O} \leq \right) \cap E_I^{\leq}$$

Umgebung von $(\mu_i)_I$. Also ist $(\mu_i^n)_I \in U$ für hinreichend

großes n , also jedenfalls für ein n . Dann gilt aber $|J^*| \leq |J|$

für jedes endliche $J^* \subset I_{\mu}$, also $|I_{\mu}| \leq |J|$, d.h.

$$(\mu_i)_I \in \{(\mu_i)_I \in E_I^{\leq} : |I_{\mu}| \leq |J|\}.$$

Aus $|J| = |I| = \infty$ folgt sogar

$$\pi^{-1}(\tilde{E}_{J \subset I}^{\leq}) = E_I^{\leq}, \text{ also } \tilde{E}_{J \subset I}^{\leq} = \tilde{E}_I^{\leq}.$$

2. inj_1 ist Isometrie bzgl. D_{∞} und D_1 und bzgl. $\mathcal{T}$, $\mathcal{T}_{\infty}$, $\mathcal{T}_1$ topologisch, wie schon oben erwähnt wurde.

Sei nun $|I| = |J| = \infty$.

Eine bijektive Abbildung $\sigma: I \rightarrow J$ induziert kanonische,

surjektive Isometrien $\sigma: E_J^{\leq} \rightarrow E_I^{\leq}$, $(\mu_j)_J \mapsto (\mu_{\sigma j})_{\sigma j \in I}$

und $\tilde{\sigma}: \tilde{E}_J^{\leq} \rightarrow \tilde{E}_I^{\leq}$, $(\tilde{\mu}_j)_J \mapsto (\tilde{\mu}_{\sigma j})_{\sigma j \in I}$ (dabei sei

$(\mu_j)_J$ ein Repräsentant von $(\tilde{\mu}_j)_J$).

Nun gilt für $(\mu_j)_J, (\mu_j)_J \in E_J^{\leq}$ $(\mu_{\sigma j})_I \sim \text{inj}(\mu_j)_J$

und $(\mu_{\sigma j})_I \sim \text{inj}(\mu_j)_J$, also für $D^* \in \{D_{\infty}, D_1\}$:

$$D^*((\tilde{\mu}_j)_J, (\tilde{\mu}_j)_J) = D^*(\tilde{\sigma}(\tilde{\mu}_j)_J, \tilde{\sigma}(\tilde{\mu}_j)_J) = D^*((\tilde{\mu}_{\sigma j})_I, (\tilde{\mu}_{\sigma j})_I)$$

$$= D^*(\text{inj}(\tilde{\mu}_j)_J, \text{inj}(\tilde{\mu}_j)_J).$$

Also ist $\widetilde{\text{inj}}$ eine wegen 1) surjektive Isometrie. Nach Konstruktion ist $\widetilde{\text{inj}}_2$ die Identität.

Sei $|J| < \infty$;

Zu $(\mu_i)_I, (\nu_i)_I \in E_I^{\leq}$ mit $|I_\mu|, |I_\nu| \leq |J| < \infty$ gibt es wegen

$$d(\mu, \nu) \leq d(\mu, 0) + d(0, \nu) \quad (\mu_i^*)_I, (\nu_i^*)_I \in E_I^{\leq} \text{ mit}$$

$$(\mu_i^*)_I \sim (\mu_i)_I, (\nu_i^*)_I \sim (\nu_i)_I \text{ und } I_{\mu^*}, I_{\nu^*} \in J \text{ und}$$

$$D^*((\mu_i)_I, (\nu_i)_I) \geq D^*((\mu_i^*)_I, (\nu_i^*)_I) \text{ für } D^* \in \{D_\infty, D_1\}.$$

Daher folgt: $\widetilde{\text{inj}}: \widetilde{E}_J^{\leq} \longrightarrow \widetilde{E}_I^{\leq}$ ist zu D^* auf $\widetilde{E}_J^{\leq}$ bzw. $\widetilde{E}_I^{\leq}$

eine Isometrie. $\widetilde{\text{inj}}_1$ ist also topologisch und $\widetilde{\text{inj}}$ wegen

1) abgeschlossen.

Bemerkung

16. Für $I = N$, $|J| < \infty$ ist $\widetilde{\text{inj}}$ bzgl. $\widetilde{\mathcal{F}}^* = \widetilde{\mathcal{F}}$ keine Einbettung. $\widetilde{\text{inj}}$ ist zwar stetig, $\widetilde{\text{inj}}(\widetilde{E}_J^{\leq})$ ist auch abgeschlossen, $\widetilde{\text{inj}}$ ist aber keine abgeschlossene Abbildung: $(\widetilde{E}_J^{\leq}, \widetilde{\mathcal{F}})$ ist Hausdorffraum, in $(\widetilde{E}_I^{\leq}, \widetilde{\mathcal{F}})$ enthält jede abgeschlossene Menge das Element $(0, 0, \dots)$.

Definition

$\overline{D}_\infty$ sei die Hausdorffmetrik auf $\widetilde{E}_I^{\leq}$ zu D_∞ auf $E_I^{\leq}$.

$\overline{D}_1$ sei die Hausdorffmetrik auf $\widetilde{E}_I^{\leq}$ zu D_1 auf $E_I^{\leq}$.

III Gütefunktionen und Clusterungen

Für den weiteren Ausbau einer Theorie der Clusteranalyse werden im folgenden einige zumeist naheliegende Begriffe benötigt. Eine Clusterungsfunktion ist eine Abbildung, die einem W-Maß $\mu \in \mathfrak{M}^1$ eine Clusterung $(\tilde{\mu}_i)_I \in \tilde{E}_I^{\leq}$ zuordnet. Es ist dabei sinnvoll, sich auf Clusterungen $(\tilde{\mu}_i)_I$ von μ mit $\Sigma(\tilde{\mu}_i)_I \leq \mu$ zu beschränken, also $(\tilde{\mu}_i)_I \in \tilde{E}_I^{\leq}(\mu)$ zu wählen. Da man daran interessiert ist, in einem gewissen Sinn optimale Clusterungen zu bilden, ordnet man durch eine geeignet zu wählende Gütefunktion verschiedenen Clusterungen einer Verteilung reelle Werte zu und wählt solche Clusterungen, die eine besonders hohe Güte erzielen. Da wir i.a. nicht nur Clusterungen $(\tilde{\mu}_i)_I$ zulassen mit der Eigenschaft, daß $\Sigma_I \mu_i$ wieder ein W-Maß ist, kann eine Clusterung zu verschiedenen W-Maßen gleichzeitig gehören. Will man die Güte einer Clusterung auch von dem W-Maß abhängig machen, das in Cluster zerlegt wird, so ist es notwendig, den Definitionsbereich von Gütefunktionen bzw. den Wertebereich von Clusterungsfunktionen nicht in $\tilde{E}_I^{\leq}$ zu wählen, sondern in der Menge $\bigcup_{\mu \in \mathfrak{M}^1} \{\mu\} \times \tilde{E}_I^{\leq}(\mu)$. Die erste Komponente eines Punktes $(\mu, (\tilde{\mu}_i)_I)$ dieser Menge bringt dabei zum Ausdruck, zu welchem W-Maß μ die Clusterung $(\tilde{\mu}_i)_I$ gehört.

Außerdem wollen wir auch Gütefunktionen untersuchen, die nicht für alle W-Maße und alle Clusterungen erklärt sind, z.B. solche, die die Existenz von Momenten des W-Maßes $\mu \in \mathfrak{M}^1$ voraussetzen. Deshalb betrachten wir nur einen Teilbereich $\mathfrak{D}$ von W-Maßen und gewisse Teilsysteme $A(\mathfrak{D})$ im Raum $\bigcup_{\mu \in \mathfrak{M}^1} \{\mu\} \times \tilde{E}_I^{\leq}(\mu)$.

Die Auffassung von Clusteranalyse als einem Teilgebiet der

Statistik ist der wichtigste Gesichtspunkt für die Wahl des im II. Kapitel beschriebenen Modells. Clusterungen empirischer Maße sollen als Schätzungen der Clusterung des dem Versuch zugrundeliegenden Wahrscheinlichkeitsmaßes dienen. Soll diese Auffassung gerechtfertigt sein, so müssen die Clusterungen der empirischen Maße die Clusterung des zugrundeliegenden Maßes immer besser approximieren, je größer der Stichprobenumfang wird, bzw. je besser das empirische Maß das zugrundeliegende Maß approximiert; d.h. die Clustermethode soll konsistent sein. Eine solche rein asymptotische Forderung reicht aber besonders in den Fällen nicht aus, in denen man eine Clusteranalyse wünscht, wenn nämlich ein umfangreiches Datenmaterial vorliegt, aber nur sehr unzureichende Modellvorstellungen der jeweiligen Fachwissenschaft vorliegen. Clusteranalyse dient dann ja den Zwecken einer deskriptiven Statistik. Statt der Konsistenz untersucht man daher besser die Robustheit, d.h. Stetigkeit. (Vgl. Huber (1975, b).) Da Clusterungsfunktionen in der Literatur meist implizit über Gütefunktionen definiert sind, befassen sich die folgenden Untersuchungen somit hauptsächlich mit der Stetigkeit von Gütefunktionen und Clusterungsfunktionen oder allgemeiner mit den topologischen Beziehungen zwischen Gütefunktionen und Clusterungsfunktionen.

Die Stetigkeitsforderung ist noch aus einem weiteren Grunde wichtig. Überträgt man nämlich die Modellvorstellungen dieser Arbeit auf die in der Literatur beschriebenen Gütefunktionen und Clusterungsfunktionen, so sind diese zuerst nur auf empirischen Maßen bzw. auf mengenartigen Clusterungen definiert. Durch die topologische Betrachtung ist es jedoch häufig möglich, sie geeignet fortzusetzen, da empirische

Maße in $\mathcal{M}^1$ bzw. mengenartige Clusterungen in $\tilde{E}_I^{\leq}$ dicht liegen. (Vgl. I Satz 2 und II Satz 6.)

Sei $\mathcal{D} \subset \mathcal{M}^1$ mit der schwachen Topologie versehen. Der Raum $\mathcal{D} \times \tilde{E}(\mathcal{D})$ trage die von der Summenmetrik

$$d_A := d + \overline{D}_\infty := d \circ (\text{pr}_1 \times \text{pr}_1) + \overline{D}_\infty (\text{pr}_2 \times \text{pr}_2)$$

erzeugte Produkttopologie zu den Topologien $w(X)$ auf $\mathcal{D}$ und $\tilde{\tau}_\infty$ auf $\tilde{E}(\mathcal{D})$.

Definition

$$A_0(\mathcal{D}) := \bigcup_{\mu \in \mathcal{D}} \{\mu\} \times \tilde{E}(\mu) .$$

Bemerkungen

1. Für $(\mu, (\tilde{\mu}_i)_I) \in A_0(\mathcal{D})$ gilt: $\Sigma(\tilde{\mu}_i)_I \leq \mu$.

2. $A_0(\mathcal{D})$ ist abgeschlossen in $\mathcal{D} \times \tilde{E}(\mathcal{D})$.

Vgl. II Satz 6.

Definition

Eine Menge $A(\mathcal{D})$ heißt abgeschlossenes genügend dichtes System in $A_0(\mathcal{D})$, wenn die folgenden Bedingungen gelten:

- (1) $A(\mathcal{D})$ ist eine abgeschlossene Teilmenge von $A_0(\mathcal{D})$.
- (2) Zu jedem $\mu \in \mathcal{D}$ gibt es ein $(\mu, (\tilde{\mu}_i)_I) \in A(\mathcal{D})$.
- (3) Zu jeder gegen $\mu \in \mathcal{D}$ konvergenten Folge $(\mu^n)_{n \in \mathbb{N}}$ in $\mathcal{D}$ und jeder Clusterung $(\tilde{\mu}_i)_I$ mit $(\mu, (\tilde{\mu}_i)_I) \in A(\mathcal{D})$ gibt es eine Folge $(\mu^n, (\tilde{\mu}_i^n)_I)_{n \in \mathbb{N}}$ in $A(\mathcal{D})$, die gegen $(\mu, (\tilde{\mu}_i)_I)$ konvergiert.

Ferner seien für $\mu \in \mathcal{D}$

$$\begin{aligned} A(\mu) &:= \{(\gamma, (\tilde{\gamma}_i)_I) \in A(\mathcal{D}) : \text{pr}_1(\gamma, (\tilde{\gamma}_i)_I) = \mu\} \\ &= A(\mathcal{D}) \cap (\{\mu\} \times \tilde{E}_I^{\leq}(\mu)) \end{aligned}$$

und für $\mathcal{K} \subset \mathcal{D}$

$$\begin{aligned} A(\mathcal{K}) &:= \{(\gamma, (\tilde{\gamma}_i)_I) \in A(\mathcal{D}) : \text{pr}_1(\gamma, (\tilde{\gamma}_i)_I) \in \mathcal{K}\} \\ &= A(\mathcal{D}) \cap (\mathcal{K} \times \tilde{E}_I^{\leq}(\mathcal{K})) \end{aligned}$$

definiert.

Bemerkungen

1. Mit $\mathcal{K} = \mathcal{D}$ ist auch $A(\mathcal{K})$ kompakt.

Denn $A(\mathcal{D})$ ist abgeschlossen in $A_0(\mathcal{D})$ und damit in $\mathcal{D} \times \tilde{E}(\mathcal{D})$. Also ist $A(\mathcal{K})$ abgeschlossen in der kompakten Menge $\mathcal{K} \times \tilde{E}(\mathcal{K})$, also kompakt.

2. $A(\mathcal{D}) = \bigcup_{\mu \in \mathcal{D}} A(\mu)$ mit $A(\mu) \cap A(\nu) = \emptyset$ für $\mu \neq \nu$.

Beispiele für abgeschlossene und genügend dichte Systeme.

$$A_0(\mathcal{D})$$

$$A_1(\mathcal{D}) := \{(\mu, (\tilde{\mu}_i)_I) \in A_0(\mathcal{D}) : |I_\mu| \leq j\} \quad \text{für ein } j \in \mathbb{N}.$$

$$A_2(\mathcal{D}) := \{(\mu, (\tilde{\mu}_i)_I) \in A_0(\mathcal{D}) : \Sigma(\tilde{\mu}_i)_I(X) = 1\} \quad \text{für } |I| < \infty.$$

$$A_3(\mathcal{D}) := \{(\mu, (\tilde{\mu}_i)_I) \in A_0(\mathcal{D}) : 0 \leq \varepsilon \leq \mu_i(X) \quad \text{für } i \in I\}$$

für endliches I und $|I| \leq 1$.

$$A_4(\mathcal{D}) := A_2(\mathcal{D}) \cap A_3(\mathcal{D}).$$

$$A_5(\mathcal{D}) := \{(\mu, (\tilde{\mu}_i)_I) \in A_0(\mathcal{D}) : \Sigma(\tilde{\mu}_i)_I \leq f\mu\}$$

für $f \in C(X)$, $0 \leq f \leq 1$.

$$A_6(\mathcal{D}) := \{(\mu^n, (\tilde{\mu}_i^n)_I) \in A_0(\mathcal{D}) : n \in \mathbb{N} \cup \{0\}\}, \text{ wenn } \mu^n \neq \mu^0 \text{ für } n \in \mathbb{N}$$

und $(\mu^n, (\tilde{\mu}_i^n)_I) \xrightarrow{n \rightarrow \infty} (\mu^0, (\tilde{\mu}_i^0)_I) \quad \text{in } A_0(\mathcal{D}).$

$$A_7(\mathcal{D}) := \phi(\mathcal{D}) \quad \text{für stetige Clusterungsfunktionen } \phi.$$

Vgl. Satz 4 dieses Kapitels weiter unten.

Zum Nachweis der Eigenschaft (3) vgl. I Satz 2 und II Satz 6.

Zum System $A_1(\mathcal{D})$ vergleiche auch Satz II, 7.

Definition

Eine Abbildung $\phi: \mathcal{D} \longrightarrow A(\mathcal{D})$ heißt Clusterungsfunktion, wenn gilt:

$$\text{pr}_1 \circ \phi = \text{id}_{\mathcal{D}} \quad (\text{oder } \phi = (\text{id}_{\mathcal{D}}, \text{pr}_2 \circ \phi)).$$

Gilt $\phi(\mu) = (\mu, (\tilde{\mu}_1)_I)$, so ist $(\tilde{\mu}_1)_I$ die Clusterung von μ unter ϕ .

ϕ heißt Clusterungsfunktion ohne Masseverlust, wenn zusätzlich gilt:

$$\Sigma \circ \text{pr}_2 \circ \phi = \text{id}_{\mathcal{D}}.$$

Für die Clusterung $(\tilde{\mu}_1)_I$ von μ unter ϕ gilt dann:

$$\Sigma(\tilde{\mu}_1)_I = \mu.$$

Die Clusterungen aller Maße aus $\mathcal{D}$ liegen dann in $\tilde{E}_I^1$.

Definition

1. $G : A(\mathcal{D}) \longrightarrow \mathbb{R}$ sei eine Abbildung.

$M_G : \mathcal{D} \longrightarrow \bar{\mathbb{R}}$ sei dann definiert durch

$$M_G(\mu) := \sup G(A(\mu)).$$

$\hat{\phi}_G : \mathcal{D} \longrightarrow \mathcal{P}(A(\mathcal{D}))$ sei definiert durch

$$\begin{aligned} \hat{\phi}_G(\mu) &:= \{ (\mu, (\tilde{\mu}_1)_I) \in A(\mu) : G(\mu, (\tilde{\mu}_1)_I) = M_G(\mu) \} \\ &= A(\mu) \wedge G^{-1}(M_G(\mu)). \end{aligned}$$

Weiterhin sei für eine Menge $\mathcal{K} \subseteq \mathcal{D}$

$$\hat{\phi}_G(\mathcal{K}) := \bigcup_{\mu \in \mathcal{K}} \hat{\phi}_G(\mu).$$

2. G heißt Gütefunktion, wenn für jedes $\mu \in \mathcal{D}$ die Menge $\hat{\phi}_G(\mu)$ nicht leer ist. Damit gleichwertig ist: Für $\mu \in \mathcal{D}$ gilt $M_G(\mu) \in G(A(\mu))$.

3. Eine Gütefunktion G heißt streng in μ , wenn $|\hat{\phi}_G(\mu)| = 1$ gilt. G heißt streng, wenn G für jedes $\mu \in \mathcal{D}$ streng in μ ist.

4. Sei $G : A(\mathcal{D}) \longrightarrow \mathbb{R}$ eine Gütefunktion.

Eine Clusterungsfunktion $\phi : \mathcal{D} \longrightarrow A(\mathcal{D})$ heißt passend zu G , wenn $G \circ \phi = M_G$ gilt. Damit gleichwertig ist $\phi(\mu) \in \hat{\phi}_G(\mu)$ für alle $\mu \in \mathcal{D}$.

5. Die Funktion $\hat{\phi}_G$ zu einer Gütefunktion G heißt verallgemeinerte Clusterungsfunktion, das zugehörige M_G heißt maximale Güte.

Bemerkung

$G: A(\mathcal{D}) \longrightarrow \mathbb{R}$ sei eine Abbildung derart, daß für jedes $\mu \in \mathcal{D}$ gilt: $G|_{A(\mu)}$ ist nach oben halbstetig. Dann ist G eine Gütefunktion, da G auf $A(\mu)$ ihr Maximum annimmt.

Ist G eine Gütefunktion und ϕ_G eine zu G passende Clusterungsfunktion und sind beide Funktionen stetig, so auch $M_G = G \circ \phi_G$.

Es gilt aber allgemeiner der folgende Satz:

Satz 1

Sei $\mu \in \mathcal{D}$ und $G: A(\mathcal{D}) \longrightarrow \mathbb{R}$ eine Gütefunktion, so gilt:

- a) Ist G in einem Punkte $(\mu, (\tilde{\mu}_i)_I) \in \hat{\phi}_G(\mu)$ nach unten halbstetig, so ist auch M_G in μ nach unten halbstetig.
- b) Ist G in allen Punkten aus $A(\mu) \cap (\hat{\phi}_G(\mathcal{D}))^a$ nach oben halbstetig, so ist auch M_G in μ nach oben halbstetig.

Zusammen gilt also auch:

- c) Ist G in allen Punkten aus $A(\mu) \cap (\hat{\phi}_G(\mathcal{D}))^a$ stetig, so ist auch M_G in μ stetig.

Beweis:

- a) $(\mu^n)_{n \in \mathbb{N}}$ sei eine gegen $\mu \in \mathcal{D}$ konvergente Folge in $\mathcal{D}$.

Nach Definition von $A(\mathcal{D})$ existiert dann eine Folge

$((\tilde{\mu}_i^n)_I)_{n \in \mathbb{N}}$, die gegen $(\tilde{\mu}_i)_I$ konvergiert mit

$(\mu^n, (\tilde{\mu}_i^n)_I) \in A(\mu^n)$ für $n \in \mathbb{N}$. Da G in $(\mu, (\tilde{\mu}_i)_I)$ nach unten halbstetig ist, gilt:

$$\liminf_{n \rightarrow \infty} G(\mu^n, (\tilde{\mu}_i^n)_I) \geq G(\mu, (\tilde{\mu}_i)_I) = M_G(\mu).$$

Wegen $M_G(\mu^n) \geq G(\mu^n, (\tilde{\mu}_i^n)_I)$ folgt daraus:

$$\liminf_{n \rightarrow \infty} M_G(\mu^n) \geq M_G(\mu) \quad , \text{ d.h.}$$

M_G ist nach unten halbstetig in μ .

b) Da G eine Gütefunktion ist, existiert eine zu G passende Clusterungsfunktion $\phi: \mathcal{D} \rightarrow A(\mathcal{D})$. Es gelte

$\mathcal{D} \ni \mu^n \rightarrow \mu$ für $n \rightarrow \infty$. Die Menge $A(\{\mu, \mu^1, \dots\})$

ist kompakt, also besitzt jede Teilfolge $(\phi(\mu^k))_{k \in \mathbb{N}}$ der

Folge $(\phi(\mu^n))_{n \in \mathbb{N}}$ eine gegen ein $(\gamma, (\tilde{\gamma}_i)_I) \in A(\mathcal{D})$

konvergente Teil-Folge $(\phi(\mu^l))_{l \in \mathbb{N}}$. Wegen $\mu^n \rightarrow \mu$

für $n \rightarrow \infty$ ist $\gamma = \mu$ und daher $(\gamma, (\tilde{\gamma}_i)_I) \in A(\mu) \cap (\hat{\phi}_G(\mathcal{D}))^a$.

Weiter gilt:

$$\begin{aligned} \limsup_{l \rightarrow \infty} M_G(\mu^l) &= \limsup_{l \rightarrow \infty} G(\phi(\mu^l)) && (\phi \text{ ist passend zu } G) \\ &\leq G(\gamma, (\tilde{\gamma}_i)_I) && (\text{Halbstetigkeit von } G) \\ &\leq M_G(\mu) . \end{aligned}$$

Weil jede Teilfolge von $(\phi(\mu^n))_{n \in \mathbb{N}}$ eine Teil-Folge

$(\phi(\mu^l))_{l \in \mathbb{N}}$ mit $\limsup_{l \rightarrow \infty} M_G(\mu^l) \leq M_G(\mu)$ enthält, gilt

$$\limsup_{n \rightarrow \infty} M_G(\mu^n) \leq M_G(\mu) .$$

M_G ist in μ nach oben halbstetig.

c) folgt sofort aus a) und b).

Q.e.d.

Bei der Betrachtung konkreter Beispiele erweist sich der Begriff des abgeschlossenen genügend dichten Systems als sehr nützlich, da man sich damit auf geeignete Definitionsbereiche einschränken kann. Für die theoretische Untersuchung stetiger Gütefunktionen kann als $A(\mathcal{D})$ aber immer $A_0(\mathcal{D})$ angenommen werden. Dies ergibt sich aus dem folgenden Satz.

Satz 2

Jede beschränkte, stetige Gütefunktion G auf $A(\mathcal{D})$ läßt sich so zu einer stetigen Gütefunktion $\bar{G}$ auf $A_0(\mathcal{D})$ fortsetzen, daß $\hat{\phi}_{\bar{G}} = \hat{\phi}_G$ ist.

Bemerkung

Der Übergang von G zu $\arctg \circ G$ zeigt, daß die Voraussetzung der Beschränktheit von G unwesentlich ist.

Beweis:

Mit G ist M_G beschränkt, daher gibt es eine untere Schranke $-K$ für $G - M_G$.

Die Funktion $G - M_G: A(\mathcal{D}) \longrightarrow [-K, 0]$

$$(\mu, (\tilde{\mu}_1)_I) \longmapsto G(\mu, (\tilde{\mu}_1)_I) - M_G(\mu)$$

läßt sich nach dem Satz von Tietze stetig auf $A_0(\mathcal{D})$ fortsetzen. Die Fortsetzung G' ist dabei so wählbar, daß sie keine positiven Werte annimmt. Sei

$$\bar{G} := G' + M_G - d_A(\cdot, A(\mathcal{D})) \quad \text{definiert durch}$$

$$\bar{G}(\mu, (\tilde{\mu}_1)_I) := G'(\mu, (\tilde{\mu}_1)_I) + M_G(\mu) - d_A((\mu, (\tilde{\mu}_1)_I), A(\mathcal{D})).$$

Auf $A(\mathcal{D})$ gilt $\bar{G} = G' + M_G = G - M_G + M_G = G$.

Auf dem Komplement von $A(\mathcal{D})$ gilt $\bar{G} < M_G \circ \text{pr}_1$,

da $A(\mathcal{D})$ in $A_0(\mathcal{D})$ abgeschlossen ist. Also gilt $\hat{\phi}_{\bar{G}} \subset \hat{\phi}_G$.

Da auf $A(\mathcal{D})$ aber $\bar{G} = G$ ist, folgt auch $\hat{\phi}_{\bar{G}} \supset \hat{\phi}_G$. Q.e.d.

Für eine Menge $B \subset A(\mathcal{D})$ und $\varepsilon > 0$ sei wie üblich

$$U^\varepsilon(B) := \{(\mu, (\tilde{\mu}_1)_I) \in A(\mathcal{D}) : d_A((\mu, (\tilde{\mu}_1)_I), B) < \varepsilon\}$$

die ε -Umgebung von B in $A(\mathcal{D})$. Der folgende Satz zeigt,

daß sich die Clusterungen in $\hat{\phi}_G(\mu^n)$ den Clusterungen

in $\hat{\phi}_G(\mu)$ beliebig genau nähern, wenn $(\mu^n)_{n \in \mathbb{N}}$ gegen μ

konvergiert. Da er eine Rechtfertigung für die Cluster-

methoden liefert, die über Gütefunktionen implizit definiert werden, ist er für die Anwendung der wichtigste Satz dieser Arbeit.

Satz 3

Sei $\mu \in \mathcal{D}$ und G eine in allen Punkten von $A(\mu) \cap (\hat{\Phi}_G(\mathcal{D}))^a$ stetige Gütefunktion, $(\mu^n)_{n \in \mathbb{N}}$ konvergiere in $\mathcal{D}$ gegen μ .

Dann gilt:

a) Für jedes $\varepsilon > 0$ existiert ein $n_0 \in \mathbb{N}$, so daß für alle $n \geq n_0$

$$\hat{\Phi}_G(\mu^n) \subset U^\varepsilon(\hat{\Phi}_G(\mu)) .$$

b) Insbesondere gilt für jede zu G passende Clusterungsfunktion ϕ und für jedes $\varepsilon > 0$

$$\phi(\mu^n) \in U^\varepsilon(\hat{\Phi}_G(\mu)) \quad \text{für genügend großes } n .$$

c) Ist G in μ streng, so gilt für jede zu G passende Clusterungsfunktion ϕ

$$\phi(\mu^n) \longrightarrow \phi(\mu) \quad \text{für } n \longrightarrow \infty .$$

ϕ ist also in μ stetig.

Beweis:

a) Wir nehmen für einen indirekten Beweis an, daß für ein festes $\varepsilon > 0$ immer wieder $\hat{\Phi}_G(\mu^n) \not\subset U^\varepsilon(\hat{\Phi}_G(\mu))$ gilt.

Dann existieren eine Teilfolge $(\hat{\Phi}_G(\mu^k))_{k \in \mathbb{N}}$ und

$$(\mu^k, (\tilde{\mu}_i^k)_I) \in \hat{\Phi}_G(\mu^k) \quad \text{mit} \quad (\mu^k, (\tilde{\mu}_i^k)_I) \notin U^\varepsilon(\hat{\Phi}_G(\mu)) .$$

Da $A(\{\mu, \mu^1, \mu^2, \dots\})$ kompakt ist, ergibt sich daraus

wie in Satz 1, b) die Konvergenz einer Teilfolge $(\mu^1, (\tilde{\mu}_i^1)_I)_{1 \in \mathbb{N}}$

der Folge $(\mu^k, (\tilde{\mu}_i^k)_I)_{k \in \mathbb{N}}$ gegen ein Element

$$(\mu, (\tilde{\mu}_i)_I) \in A(\mu) \cap (\hat{\Phi}_G(\mathcal{D}))^a .$$

Wegen der Stetigkeitsvoraussetzung an G gilt

$$M_G(\mu^1) = G(\mu^1, (\tilde{\mu}_i^1)_I) \longrightarrow G(\mu, (\tilde{\mu}_i)_I)$$

und, da M_G in μ wegen Satz 1 stetig ist,

$$M_G(\mu^1) \longrightarrow M_G(\mu).$$

Damit gilt $(\mu, (\tilde{\mu}_1)_I) \in \hat{\Phi}_G(\mu)$, also

$$(\mu^1, (\tilde{\mu}_1^1)_I) \in U^\varepsilon(\hat{\Phi}_G(\mu)) \quad \text{für große } 1.$$

Dies ist ein Widerspruch zur obigen Annahme, also gilt a).

b) und c) sind direkte Folgerungen von a).

Q.e.d.

Die Behauptung in a) des Satzes ist analog einer Konvergenzbedingung für Folgen aufgebaut. Ist G in μ streng,

so gilt auch $d_A(\hat{\Phi}_G(\mu^n), \hat{\Phi}_G(\mu)) \longrightarrow 0$ für $n \longrightarrow \infty$.

Es liegt daher nahe, die Aussage a) als (Folgen-)Stetigkeit der verallgemeinerten Clusterungsfunktion $\hat{\Phi}_G$ in μ zu bezeichnen. Es sei aber ausdrücklich vermerkt, daß dies keine Stetigkeit im Sinne der Hausdorff-Metrik bedeutet. $\hat{\Phi}_G(\mu)$ kann Elemente enthalten, die keine Häufungspunkte von Folgen aus $\hat{\Phi}_G(\mu^n)$ sind.

Eine ähnliche Aussage wie in Satz 3 erweist sich noch als nützlich:

Satz 3*

Für ein $\varepsilon > 0$ sei G eine in allen Punkten von

$A(\mu) \cap (G^{-1} [M_G(\mu) - \varepsilon, M_G(\mu) + \varepsilon])^a$ stetige Gütefunktion

und es gelte $\mu^n \longrightarrow \mu$.

Gilt dann für die Folge $(\mu^n, (\tilde{\mu}_1^n)_I)_N$ in $A(\mathcal{D})$

$G(\mu^n, (\tilde{\mu}_1^n)_I) \longrightarrow M_G(\mu)$, so liegen alle Häufungspunkte von $(\mu^n, (\tilde{\mu}_1^n)_I)_N$ in $\hat{\Phi}_G(\mu)$.

Beweis:

$(\mu, (\tilde{\mu}_1)_I)$ sei der Limes einer Teilfolge $(\mu^1, (\tilde{\mu}_1^1)_I)_{1 \in \mathbb{N}}$.

Dann gilt $(\mu, (\tilde{\mu}_1)_I) \in A(\mu) \cap (G^{-1} [M_G(\mu) - \varepsilon, M_G(\mu) + \varepsilon])^a$,

also ist G stetig in $(\mu, (\tilde{\mu}_i)_I)$. Wegen

$$M_G(\mu) = \lim_n G(\mu^n, (\tilde{\mu}_i^n)_I) = \lim_1 G(\mu^1, (\tilde{\mu}_i^1)_I) = G(\mu, (\tilde{\mu}_i)_I)$$

gilt $(\mu, (\tilde{\mu}_i)_I) \in \hat{\Phi}_G(\mu)$.

Q.e.d.

Für die Anwendung ist dieser Satz wichtig, da er die Konsistenz gewisser Clustermethoden liefert, die in der

Literatur vorgeschlagen werden. Denn seien

$\mu^n = \frac{1}{n} \sum_{k=1}^n \delta_{x_k}$ empirische Maße zu einer zugrundeliegenden Verteilung μ mit $\mu^n \rightarrow \mu$. Dann wählt man

unter allen mengenartigen Clusterungen $(\tilde{\mu}_i^n)_I$ von μ^n ,

das sind Clusterungen mit $\mu_i^n = \frac{1}{n} \sum_{\text{gewisse } k} \delta_{x_k}$ für

alle $i \in I$, diejenigen aus, die eine jeweils geeignete stetige

Gütefunktion $G: A_0(\mathcal{D}) \rightarrow \mathbb{R}$ maximieren und bezeichnet die

Menge dieser Clusterungen mit $\hat{\Phi}_G^n(\mu^n)$. Zu einem Element

$(\mu, (\tilde{\mu}_i)_I) \in \hat{\Phi}_G(\mu)$ existiert wegen II Satz 6 eine Folge

mengenartiger Clusterungen $((\tilde{\mu}_i^n)_I)_{n \in \mathbb{N}}$ mit $(\tilde{\mu}_i^n)_I \in A(\mu^n)$

und $(\tilde{\mu}_i^n)_I \rightarrow (\tilde{\mu}_i)_I$. Also gilt wegen der Stetigkeit

von G auch $G(\mu^n, (\tilde{\mu}_i^n)_I) \rightarrow M_G(\mu)$.

Für Elemente $(\mu^n, (\tilde{\mu}_i^n)_I) \in \hat{\Phi}_G^n(\mu^n)$ gilt aber

$G(\mu^n, (\tilde{\mu}_i^n)_I) \geq G(\mu^n, (\tilde{\mu}_i^n)_I)$, also

$G(\mu^n, (\tilde{\mu}_i^n)_I) \rightarrow M_G(\mu)$.

Wegen Satz 3* liegen also alle Häufungspunkte der Folge

$(\mu^n, (\tilde{\mu}_i^n)_I)$ in $\hat{\Phi}_G(\mu)$. Ist G streng in μ , so folgt die

Behauptung der Konsistenz.

In den Sätzen 1 und 3, die sich lokal auf ein $\mu \in \mathcal{D}$

beziehen, wird die Stetigkeit der Gütefunktion in

$(\hat{\Phi}_G(\mathcal{D}))^a \cap A(\mu)$ - gewissermaßen also in der Nähe ihrer

Maximalstellen - vorausgesetzt. Der folgende Satz zeigt

unter anderem die Bedeutung der Menge $(\hat{\phi}_G(\mathcal{D}))^a$.

Satz 4

1) Für jede stetige Gütefunktion $G: A(\mathcal{D}) \longrightarrow \mathbb{R}$ ist die Menge der Maximalpunkte $\hat{\phi}_G(\mathcal{D})$ in $A(\mathcal{D})$ abgeschlossen.

2) Zu einer abgeschlossenen Menge $B \subseteq A(\mathcal{D})$ mit $B \cap A(\mu) \neq \emptyset$ für alle $\mu \in \mathcal{D}$, gibt es eine stetige Gütefunktion $G_B: A(\mathcal{D}) \longrightarrow \mathbb{R}$ mit $\hat{\phi}_{G_B}(\mathcal{D}) = B$.

3) Zu einer Clusterungsfunktion $\phi: \mathcal{D} \longrightarrow A_0(\mathcal{D})$ gibt es eine stetige Gütefunktion G_ϕ , zu der ϕ paßt und die genau in den Punkten streng ist, in denen ϕ stetig ist.

Beweis:

1) Die stetige Gütefunktion $G - M_G: A(\mathcal{D}) \longrightarrow \mathbb{R}$, definiert durch $G - M_G(\mu, (\tilde{\mu}_i)_I) = G(\mu, (\tilde{\mu}_i)_I) - M_G(\mu)$, hat dieselben Maximalstellen wie G . Es gilt $\hat{\phi}_G = \hat{\phi}_{G - M_G}$, also $\hat{\phi}_G(\mathcal{D}) = \hat{\phi}_{G - M_G}(\mathcal{D}) = (G - M_G)^{-1}(\{0\})$.

Diese Menge ist aber abgeschlossen, da $G - M_G$ stetig ist.

2) $G_B: A(\mathcal{D}) \longrightarrow \mathbb{R}$ sei definiert durch

$$G_B(\mu, (\tilde{\mu}_i)_I) := -d_A((\mu, (\tilde{\mu}_i)_I), B).$$

G_B ist stetig und es gilt: $G_B \leq 0$, $M_{G_B} = 0$.

$G_B(\mu, (\tilde{\mu}_i)_I) = 0$ gilt genau dann, wenn $(\mu, (\tilde{\mu}_i)_I) \in B$, da B abgeschlossen ist. Also $\hat{\phi}_{G_B}(\mathcal{D}) = B$.

3) Für $B = (\phi(\mathcal{D}))^a$ sei $G_\phi := G_B$, dann ist G_ϕ eine stetige Gütefunktion mit $\phi(\mu) \in B = \hat{\phi}_{G_\phi}(\mathcal{D})$, also paßt ϕ zu G_ϕ . Ist ϕ stetig in μ , so gilt $(\phi(\mathcal{D}))^a \cap A(\mu) = \{\phi(\mu)\}$.

Denn zu $(\mu, (\tilde{\mu}_i)_I) \in (\phi(\mathcal{D}))^a \cap A(\mu)$ existiert eine Folge $(\mu^n, (\tilde{\mu}_i^n)_I)_{n \in \mathbb{N}}$ in $\phi(\mathcal{D})$ mit

$(\mu^n, (\tilde{\mu}_i^n)_I) \longrightarrow (\mu, (\tilde{\mu}_i)_I)$ und wegen der Stetigkeit von ϕ in μ gilt auch

$$(\mu^n, (\tilde{\mu}_i^n)_I) = \phi(\mu^n) \longrightarrow \phi(\mu) = (\mu, (\tilde{\mu}_i)_I) .$$

Also ist G_ϕ in μ streng.

Ist ϕ in μ unstetig, so existieren zwei in $\mathcal{D}$ gegen μ konvergente Folgen $(\mu^n)_{n \in \mathbb{N}}$ und $(\nu^n)_{n \in \mathbb{N}}$ mit

$$\lim_n \phi(\mu^n) \neq \lim_n \phi(\nu^n) .$$

Es gilt aber $\lim_n \phi(\mu^n) \in (\phi(\mathcal{D}))^a \cap A(\mu)$ und

$$\lim_n \phi(\nu^n) \in (\phi(\mathcal{D}))^a \cap A(\mu) ,$$

also enthält $\hat{\phi}_{G_\phi}(\mu) = (\phi(\mathcal{D}))^a \cap A(\mu)$ mindestens zwei verschiedene Elemente. G_ϕ ist nicht streng in μ .

Q.e.d.

Im Falle einer abzählbaren Indexmenge $I = \mathbb{N}$ sind die Topologien $\tilde{\tau}_1$ und $\tilde{\tau}_\infty$ verschieden. Im folgenden sollen einige der Beziehungen zwischen den Topologien und den Ergebnissen dieses Kapitels dargestellt werden.

Analog zur Metrik d_A sei die Summenmetrik d_B auf $\mathcal{D} \times \tilde{E}(\mathcal{D})$ durch

$$d_B := d + \bar{D}_1 := d \circ (pr_1 \times pr_1) + \bar{D}_1 \circ (pr_2 \times pr_2) \quad \text{definiert.}$$

Zunächst ein Beispiel:

Beispiel

Satz 1 und 3 gelten nicht für die Topologie auf $A_0(\mathcal{D})$, die durch d_B erzeugt wird, falls $I = \mathbb{N}$ ist.

$(\mu^n)_{n \in \mathbb{N}}$ sei eine in $\mathcal{M}^1$ gegen μ konvergente Folge mit $\mu^n \neq \mu$ für $n \in \mathbb{N}$. Setze: $\mathcal{D} := \{\mu^n : n \in \mathbb{N}\} \cup \{\mu\}$ und

$$\mu_i^n := \begin{cases} n^{-1} \mu^n & \text{für } i \leq n \\ 0 & \text{sonst.} \end{cases}$$

Dann sind die Mengen $A(\mu)$ und $\{(\mu^n, (\mu_i^n)_I) : n \in \mathbb{N}\}$ disjunkt und bzgl. der von d_B erzeugten Topologie abgeschlossen. Denn, weil $\tilde{E}_I^1$ bzgl. $\overline{D}_1$ abgeschlossen ist, und wegen

$$\begin{aligned} & \{(\mu^n, (\tilde{\mu}_i^n)_I) : n \in \mathbb{N}\} = \\ & = (\{\mu, \mu^1, \mu^2, \dots\} \times \tilde{E}_I^1) \cap (\{(\mu^n, (\tilde{\mu}_i^n)_I) : n \in \mathbb{N}\} \cup \{(\mu, (\tilde{0})_I)\}) \end{aligned}$$

ist $\{(\mu^n, (\tilde{\mu}_i^n)_I) : n \in \mathbb{N}\}$ abgeschlossen.

(Die Vereinigungsmenge ist als konvergente Folge in der d_A -Topologie abgeschlossen, also erst recht in der d_B -Topologie.)

Es existiert somit eine stetige Funktion $F: A_0(\mathcal{D}) \longrightarrow [0, 1]$, die auf der abgeschlossenen Menge $A(\mu)$ konstant 0 und auf der abgeschlossenen Menge $\{(\mu^n, (\tilde{\mu}_i^n)_I) : n \in \mathbb{N}\}$ konstant 1 ist, sonst aber echt zwischen 0 und 1 liegt. Ferner existiert zu einem festen $(\mu, (\tilde{\mu}_i)_I) \in A(\mu)$ mit $(\tilde{\mu}_i)_I \neq (\tilde{0})_I$ eine stetige Funktion $H: A_0(\mathcal{D}) \longrightarrow [0, \frac{1}{2}]$ mit

$$H(\mu, (\tilde{\mu}_i)_I) = \frac{1}{2}, \quad H(\mu, (\tilde{0})_I) = 0 \quad \text{und} \quad H < \frac{1}{2} \quad \text{sonst.}$$

$G := \max(F, H)$ ist eine stetige Gütefunktion mit $M_G(\mu) = \frac{1}{2}$ und $M_G(\mu^n) = 1$ für alle $n \in \mathbb{N}$. M_G ist somit in μ nicht (nach oben halb-) stetig. Die Gütefunktion ist nach Konstruktion in μ streng. $\hat{\phi}_G$ ist in μ nicht stetig im Sinne von Satz 3.

Die Sätze dieses Kapitels liefern einen Zusammenhang zwischen d_A -stetigen Gütefunktionen und (verallgemeinerten) Clusterungsfunktionen. Von entscheidender Bedeutung für den Nachweis der Stetigkeit von M_G (Satz 1) und $\hat{\phi}_G$ bzw. ϕ_G (Satz 3) ist II, Satz 5, der nur für die Topologie $\tilde{\tau}_\infty$ gilt, aber nicht für $\tilde{\tau}_1$. (Vgl. II, Beispiel 2*).

Wie das obige Beispiel lehrt, gibt es eine in den entscheidenden Punkten d_B -stetige strenge Gütefunktion mit unstetigem M_G und d_B -unstetiger Clusterungsfunktion. Allerdings läßt sich ohne Änderung des Beweises noch die Halbstetigkeit nach unten für M_G nachweisen.

Wegen $d_B \geq d_A$ ist eine bezüglich d_A stetige Gütefunktion $G: (A(\mathcal{D}), d_A) \longrightarrow \mathbb{R}$ auch d_B -stetig und eine d_B -stetige Clusterungsfunktion $\phi: (\mathcal{D}, d) \longrightarrow (A(\mathcal{D}), d_B)$ auch d_A -stetig. Daher bleibt auch die Konstruktion von d_B -stetigen, strengen Gütefunktionen zu stetigen, und dann passenden Clusterungsfunktionen (Satz 4) trivialerweise möglich. Da die Stetigkeit der (verallgemeinerten) Clusterungsfunktionen für die inhaltliche Interpretation der Clusteranalyse entscheidend ist, wird man i.a. die Topologie $\tilde{\tau}_\infty$ bevorzugen.

Da die Topologien auf $\tilde{E}_I^1$ identisch sind, sind für Clusterungsfunktionen ohne Masseverlust d_B -Stetigkeit und d_A -Stetigkeit identisch. Da sogar die Umgebungsfilter für alle Punkte aus $\tilde{E}_I^1$ bzgl. der Topologien $\tilde{\tau}_\infty$ und $\tilde{\tau}_1$ übereinstimmen (II, Bemerkung 15) ist auch für Gütefunktionen d_B -Stetigkeit und d_A -Stetigkeit in Punkten aus $\tilde{E}_I^1$ identisch. Schwierigkeiten ergeben sich aber daraus, daß $\tilde{E}_I^1$ in $\tilde{E}_I^\leq$ nicht bzgl. $\tilde{\tau}_\infty$ abgeschlossen ist, insbesondere also kein abgeschlossenes und genügend dichtes System bildet (im Sinne der Topologie $\tilde{\tau}_\infty$). Eine d_B -stetige (= d_A -stetige) Funktion von $\tilde{E}_I^1$ in $\mathbb{R}$ läßt sich i.a. nicht stetig (bzgl. d_A) auf $\tilde{E}_I^\leq$ fortsetzen. Insgesamt ergibt sich aus diesen Überlegungen eine Rechtfertigung für die Wahl des Raumes $(\tilde{E}_I^\leq, \tilde{\tau}_\infty)$ als Raum der Clusterungen, insbesondere ist auch der Teilraum der Clusterungen ohne Masseverlust $\tilde{E}_I^1$ nicht so gut geeignet.

Zum Schluß dieses Kapitels soll anhand eines Beispiels gezeigt werden, daß eine stetige Clusterungsfunktion

$\phi: \mathcal{D} \longrightarrow A(\mathcal{D})$ i.a. nicht stetig gegen π geliftet werden kann; die Konstruktion des Quotientenraumes $\tilde{E}_I$

zu $E_I^{\leq}$ ist also i.a. nicht überflüssig. Genauer zeigen wir:

Es gibt eine stetige Clusterungsfunktion $\phi: \mathcal{D} \longrightarrow A(\mathcal{D})$ zu der keine stetige Abbildung

$$\phi^*: \mathcal{D} \longrightarrow \{(\mu, (\mu_1)_I): (\mu, (\tilde{\mu}_1)_I) \in A(\mathcal{D})\}$$

mit $\phi = (\text{id}, \pi) \circ \phi^*$ existiert.

Beispiel

Sei $I = \{1, 2\}$, $X = \{z \in \mathbb{C} : |z| = 1\}$,

$$\mathcal{D} = \left\{ \frac{1}{2}(\delta_z + \delta_{-z}) \in \mathcal{M}^1 : z \in X \right\},$$

$$A = \left\{ \left(\frac{1}{2}\delta_z, \frac{1}{2}\delta_{-z} \right) \in E_I^{\leq} : z \in X \right\} \text{ und}$$

$$\phi_2: \mathcal{D} \longrightarrow \pi(A) \text{ eine Abbildung mit } \frac{1}{2}(\delta_z + \delta_{-z}) \longmapsto \widetilde{\left(\frac{1}{2}\delta_z, \frac{1}{2}\delta_{-z} \right)}.$$

$\mathcal{D}$ ist kompakt, A kompakt und zusammenhängend.

ϕ_2 ist topologisch und damit die Clusterungsfunktion

$$\phi := (\text{id}, \phi_2): \mathcal{D} \longrightarrow A_0(\mathcal{D}) \text{ stetig.}$$

Nehmen wir nun für einen indirekten Beweis an:

Es gibt eine stetige Funktion

$$(\text{id}, \phi_2^*): \mathcal{D} \longrightarrow \{(\mu, (\mu_1)_I) : (\mu, (\tilde{\mu}_1)_I) \in A_0(\mathcal{D})\}$$

mit $\pi \circ \phi_2^* = \phi_2$. Es gilt aber mit $\sigma \in \mathcal{T}_2$, $\sigma 1 = 2$, $\sigma 2 = 1$:

$$\phi_2^*(\mathcal{D}) \cup \sigma(\phi_2^*(\mathcal{D})) = A, \text{ weil } \phi_2: \mathcal{D} \longrightarrow \pi(A) \text{ surjektiv ist,}$$

$$\phi_2^*(\mathcal{D}) \cap \sigma(\phi_2^*(\mathcal{D})) = \emptyset, \text{ weil } \phi_2^* \text{ Abbildung ist,}$$

und $\phi_2^*(\mathcal{D})$ und $\sigma(\phi_2^*(\mathcal{D}))$ sind, weil ϕ_2^* und σ stetig sind,

nicht leere, kompakte, also abgeschlossene Teilmengen von A .

A ist aber zusammenhängend. Aus diesem Widerspruch folgt die Behauptung.

IV Beispiele und Anwendungen

Intuitiv liegt es nahe, die Aufteilung eines diskreten Maßes in seine Atome für ein gutes Beispiel einer Clusterung zu halten. Die folgenden Bemerkungen sollen jedoch zeigen, daß diese Auffassung zu Schwierigkeiten führt.

Die unabhängigen Wiederholungen X_1 einer nach einem atomlosen Maß $\mu \in \mathcal{M}^1$ verteilten Zufallsgröße nehmen alle mit Wahrscheinlichkeit 1 verschiedene Werte an. Die zugehörigen empirischen Verteilungen $\mu_\omega^n := \frac{1}{n} \sum_{i=1}^n \delta_{X_1(\omega)}$ sind diskrete Maße, deren einzelne Atome mit Wahrscheinlichkeit 1 die Masse $\frac{1}{n}$ haben. Ist $I = \mathbb{N}$ und zerlegt eine Clustermethode diskrete Maße in die einzelnen Atome, so haben alle Cluster der empirischen Verteilungen zu einem atomlosen Maß mit Wahrscheinlichkeit 1 die Masse $\frac{1}{n}$ und konvergieren mit wachsendem Stichprobenumfang n gegen 0. Die Clusterungen der empirischen Maße konvergieren dann also mit Wahrscheinlichkeit 1 gegen die Clusterung, die nur aus leeren Clustern besteht.

Es ergeben sich auch schon Schwierigkeiten, wenn man $I = \{1, 2\}$ wählt und nur Mischungen von zwei Diracmaßen zuläßt. X sei ein separabler, metrisierbarer Raum, der nicht nur aus isolierten Punkten besteht. In X gelte $x_n \rightarrow x$ mit $x_n \neq x$ für $n \in \mathbb{N}$, dann folgt $\delta_{x_n} \rightarrow \delta_x$ und $\frac{1}{2}(\delta_{x_n} + \delta_x) \rightarrow \delta_x$. Für die Clusterungen, die der

Zerlegung in Atome entsprechen, gilt aber:

$$\widetilde{(\frac{1}{2} \delta_x, \frac{1}{2} \delta_{x_n})} \longrightarrow \widetilde{(\frac{1}{2} \delta_x, \frac{1}{2} \delta_x)} \neq \widetilde{(\delta_x, 0)}.$$

Diese Bemerkungen zeigen, daß die Zerlegung in Atome i.a.

keine Clustermethode mit wünschenswerten Eigenschaften liefert. Ein anderer Ansatz wird sich jedoch als nützlich erweisen.

Ist $d: X \times X \longrightarrow \mathbb{R}$ eine stetige, beschränkte Funktion auf $X \times X$, so gilt für jede Clusterung $(\tilde{\mu}_i)_I \in \tilde{E}_I^1(\mu)$ und für einen beliebigen Repräsentanten $(\mu_i)_I$ von $(\tilde{\mu}_i)_I$:

$$\begin{aligned} \iint d(x, y) \, d\mu(x) \, d\mu(y) &= \\ &= \sum_{i \in I} \iint d(x, y) \, d\mu_i(x) \, d\mu_i(y) + \sum_{\substack{i, j \in I \\ i \neq j}} \iint d(x, y) \, d\mu_i(x) \, d\mu_j(y). \end{aligned}$$

Bezeichnet $d(x, y)$ eine Art Abweichung zwischen den Punkten x und y , so kann man $\iint d(x, y) \, d\mu_i(x) \, d\mu_i(y)$ als innere Abweichung des Clusters μ_i und

$\iint d(x, y) \, d\mu_i(x) \, d\mu_j(y)$ als Abweichung zwischen den Clustern μ_i und μ_j bezeichnen. Die Gleichung besagt, daß die Summe aller inneren Abweichungen und aller Abweichungen zwischen Clustern konstant bleibt für alle Clusterungen $(\tilde{\mu}_i)_I$ von μ ohne Massenverlust. Es liegt dann nahe, als Güte einer Clusterung einen Wert zu wählen, der - in einem genauer zu präzisierenden Sinne - die inneren Abweichungen widerspiegelt, die man natürlich zu minimieren sucht. Ist d nicht negativ und gilt $d(x, x) = 0$ für alle $x \in X$, so gilt für Maße der Gestalt $\alpha \delta_z$ mit $\alpha \in (0, 1]$ und $z \in X$:

$$\iint d(x, y) \, d\alpha \delta_z(x) \, d\alpha \delta_z(y) = \alpha^2 d(z, z) = 0.$$

Die Summe der inneren Abweichungen $\sum_{i \in I} \iint d(x, y) \, d\mu_i(x) \, d\mu_i(y)$ einer Clusterung $(\tilde{\mu}_i)_I$ von μ nimmt also bei diskreten Maßen μ - insbesondere für empirische Maße - ihr Minimum 0 an, wenn alle Cluster μ_i , $i \in I$, von der Gestalt $\alpha \delta_z$ mit $z \in X$ sind. Im Falle $I = \mathbb{N}$ führt die Minimierung der Summe der inneren Abweichungen also zur Zerlegung eines diskreten

Maßes in seine Atome. Versucht man wegen der oben aufgezählten Probleme diesen Ansatz zu modifizieren, so liegt es nahe, die Güte einer Clusterung über die "Güte" der einzelnen Cluster zu bestimmen oder über die Abweichungen zwischen den Clustern. Hier soll nur noch der erste Ansatz weiter verfolgt werden. Es erweist sich dabei als günstig, statt innerer Abweichungen, die man minimiert, besser Güten zu wählen, die maximiert werden. Ist $g: \mathcal{M}^{\leq} \longrightarrow [0, \infty)$ eine Funktion, die einem Teilmaß eine positive "Güte" zuordnet, so wird man versuchen, eine Gütefunktion $G: \tilde{\mathcal{E}}_I \longrightarrow \mathbb{R}$ durch $G(\tilde{\mu}_I) := \sum_{i \in I} g(\mu_i)$ zu definieren. Die Frage, wann ein so definiertes G eine stetige Gütefunktion ist, wird weiter unten untersucht.

Bei der Wahl einer Funktion g liegt es im ersten Ansatz nahe, g nur von der "Form" eines Clusters μ_i abhängig zu machen, aber nicht von dessen Masse, d.h. g mit $g(\nu) = g(\alpha \nu)$ für beliebiges $\nu \in \mathcal{M}^1$ und $\alpha \in (0, 1]$ zu wählen. Dies würde jedoch dazu führen, daß z.B. die Güte einer Clusterung wächst, wenn man einzelne Cluster ν in zwei Teile $\frac{1}{2} \nu$ aufteilt.

Da die Güte $g(\nu)$ eines Teilmaßes ν also von der "Form" und der Masse $\nu(X)$ abhängen muß, ist der folgende Ansatz sinnvoll (und erfolgreich). Es gelte $g(\nu) = \varphi(\nu(X)) \gamma(\frac{\nu}{\nu(X)})$ für jedes $\nu \in \mathcal{M}^{\leq} \setminus \{0\}$, wobei die Funktion $\varphi: [0, 1] \longrightarrow \mathbb{R}$ den Einfluß der Masse widerspiegelt und $\gamma: \mathcal{M}^1 \longrightarrow \mathbb{R}$ den Einfluß der Form. Da $g(\nu)$ eine "Güte" des Teilmaßes ν bezeichnen soll, wird $\gamma(\delta_z)$ maximal sein für Diracmaße δ_z , d.h. Diracmaße sollen optimale Form besitzen. Sinnvollerweise muß aber $\gamma(\delta_z) < \infty$ sein für $z \in X$, da die Güte

einer Clusterung sonst nur von einem Cluster der Clusterung abhängig wäre. Ist d eine beschränkte, stetige Metrik auf X , so ist etwa $\mathcal{V}$ mit

$$\mathcal{V}(\mu) := \frac{1}{1 + \iint d(x,y) d\mu(x) d\mu(y)} \quad \text{für } \mu \in \mathcal{M}^1$$

geeignet; es gilt $\mathcal{V} : \mathcal{M}^1 \rightarrow [0, \infty)$ ist stetig und beschränkt.

Damit in einer Clusterung nicht zwei Cluster $\mu_1, \mu_2 \neq 0$ mit $\mu_1 = \alpha_1 \mu$, $\mu_2 = \alpha_2 \mu$ und $\mu \in \mathcal{M}^1$ vorkommen können - die Cluster wären bis auf die Masse identisch -, wird man $\mathcal{V}$ streng superadditiv wählen; mit $\mathcal{V}(\mu) \neq 0$ gilt dann nämlich:

$$\begin{aligned} \mathcal{E}(\mu_1) + \mathcal{E}(\mu_2) &= \mathcal{V}(\alpha_1) \mathcal{V}(\mu) + \mathcal{V}(\alpha_2) \mathcal{V}(\mu) = \\ &= (\mathcal{V}(\alpha_1) + \mathcal{V}(\alpha_2)) \mathcal{V}(\mu) < \mathcal{V}(\alpha_1 + \alpha_2) \mathcal{V}(\mu) = \\ &= \mathcal{E}(\mu_1 + \mu_2). \end{aligned}$$

Ist $\mathcal{V}$ auf $(0, 1]$ zusätzlich positiv, so ist $\mathcal{V}$ streng isoton. Für $\mathcal{V}$ eignen sich beispielsweise die Funktionen $x \mapsto x^t$ für $t > 1$ oder $x \mapsto \exp(-x^{-2})$.

Um die Stetigkeit von G zu erzwingen, wird man von $\mathcal{V}$ weitere Eigenschaften verlangen: An der Stelle 0 muß $\mathcal{V}$ schneller als von erster Ordnung verschwinden. Der nächste Satz zeigt nun, daß ein solcher Ansatz sinnvoll ist. Dazu benötigt man noch folgende

Definition

Für $\mathcal{D} \subset \mathcal{M}^1$ sei

$$\begin{aligned} \mathcal{D}^* &:= \{ \mu \in \mathcal{M}^1 : \alpha \mu \in \mathcal{D} \text{ für ein } \alpha \in (0, 1] \} \\ &= \{ \mu \in \mathcal{M}^1 : \mu = \frac{\nu}{\nu(X)} \text{ für ein } \nu \neq 0, \nu \leq \lambda \in \mathcal{D} \}. \end{aligned}$$

Satz 1

1) $\varphi : [0,1] \longrightarrow \mathbb{R}$ sei eine stetige Funktion mit $\varphi(0) = 0$.
Für ein $\mathcal{D} \subset \mathcal{M}^1$ sei $\gamma : \mathcal{D}^* \longrightarrow \mathbb{R}$ eine stetige und beschränkte Funktion.

Dann ist $g : \mathcal{D}^{\leq} \longrightarrow \mathbb{R}$ definiert durch

$$g(\mu) := \begin{cases} \varphi(\mu(X)) \gamma\left(\frac{\mu}{\mu(X)}\right) & \text{für } \mu \neq 0 \\ 0 & \text{für } \mu = 0 \end{cases}$$

stetig.

2) Ist I endlich oder

gilt im Falle $I = \mathbb{N}$ zusätzlich $\lim_{\alpha \searrow 0} \frac{\varphi(\alpha)}{\alpha} = 0$,

so ist $G : A_0(\mathcal{D}) \longrightarrow \mathbb{R}$ definiert durch

$$G(\mu, (\tilde{\mu}_i)_I) := \sum_{i \in I} g(\mu_i)$$

eine stetige Gütefunktion.

Beweis:

1) Die Behauptung ergibt sich sofort aus den Voraussetzungen.

2) Für endliches I folgt die Behauptung sofort aus 1) und der Invarianz von G unter $\mathcal{T}(I)$.

Für $I = \mathbb{N}$ folgt aus der Voraussetzung $\lim_{\alpha \searrow 0} \frac{\varphi(\alpha)}{\alpha} = 0$

und der Stetigkeit der Funktion $\alpha \longmapsto \frac{\varphi(\alpha)}{\alpha}$ in $(0,1]$

die Existenz einer Schranke $K_\varphi < \infty$ mit $|\alpha^{-1} \varphi(\alpha)| \leq K_\varphi$

für $\alpha \in (0,1]$. Ist K_γ eine Schranke für γ , so ist wegen

$$\begin{aligned} \sum_I |g(\mu_i)| &= \sum_{i \in I, \mu_i \neq 0} \left| \varphi(\mu_i(X)) \gamma\left(\frac{\mu_i}{\mu_i(X)}\right) \right| \leq \\ &\leq \sum_I K_\varphi \mu_i(X) K_\gamma \leq K_\varphi K_\gamma \end{aligned}$$

G eindeutig definiert und $|G| \leq K_\varphi K_\gamma$.

Konvergiert eine Folge $(\mu^n, (\tilde{\mu}_i^n)_I)_{n \in \mathbb{N}}$ in $A_0(\mathcal{D})$ gegen

$(\mu, (\tilde{\mu}_i)_I) \in A_0(\mathcal{D})$, so gilt $(\tilde{\mu}_i^n)_I \longrightarrow (\tilde{\mu}_i)_I$ bzgl. $\tilde{\tau}_\infty$
für $n \longrightarrow \infty$.

Nach II Satz 3 gibt es geeignete Repräsentanten $(\mu_i^n)_I$ von $(\tilde{\mu}_i^n)_I$ und $(\mu_i)_I$ von $(\tilde{\mu}_i)_I$ mit

$$* \quad (\mu_i^n)_I \longrightarrow (\mu_i)_I \quad \text{bzgl. } \tilde{\tau}_\infty \text{ für } n \longrightarrow \infty.$$

Wegen $\lim_{\alpha \searrow 0} \frac{\varphi(\alpha)}{\alpha} = 0$ existiert für jedes $\eta > 0$ ein $\varepsilon > 0$,

sodaß für alle ε^* mit $0 < \varepsilon^* \leq \varepsilon$ gilt $0 \leq \frac{\varphi(\varepsilon^*)}{\varepsilon^*} < \eta$.

Zu einem beliebigen vorgegebenen $\eta > 0$ sei ein solches

$\varepsilon > 0$ gewählt, dann gilt für genügend großes n wegen *

auch $\mu_i^n(x) < \varepsilon$ für alle $i \in I$ mit $\mu_i(x) < \varepsilon/2$.

Es gilt weiter

$$\begin{aligned} & \left| G(\mu^n, (\tilde{\mu}_i^n)_I) - G(\mu, (\tilde{\mu}_i)_I) \right| = \left| \sum_{i \in I} (g(\mu_i^n) - g(\mu_i)) \right| \leq \\ & \leq \sum_{i, \mu_i(x) \geq \varepsilon/2} |g(\mu_i^n) - g(\mu_i)| + \sum_{i, \mu_i(x) < \varepsilon/2} |g(\mu_i^n) - g(\mu_i)| \end{aligned}$$

Da die erste Summe nur endlich viele Summanden enthält und

g stetig ist, konvergiert sie gegen 0 für $n \longrightarrow \infty$. Für die

zweite Summe gilt dann, wenn man der Einfachheit halber

$\varphi\left(\frac{\mu}{\mu(x)}\right) = 0$ setzt für $\mu = 0$:

$$\begin{aligned} & \sum_{i, \mu_i(x) < \varepsilon/2} |g(\mu_i^n) - g(\mu_i)| \leq \\ & \leq \sum_{i, \mu_i(x) < \varepsilon/2} \left| \varphi\left(\frac{\mu_i^n(x)}{\mu_i^n(x)}\right) \right| \left| \varphi\left(\frac{\mu_i^n}{\mu_i^n(x)}\right) \right| + \left| \varphi(\mu_i(x)) \right| \left| \varphi\left(\frac{\mu_i}{\mu_i(x)}\right) \right| \leq \\ & \leq K \sum_{i, \mu_i(x) < \varepsilon/2} \eta \mu_i^n(x) + \eta \mu_i(x) \leq \\ & \leq 2 K \eta \quad \text{für genügend großes } n. \end{aligned}$$

Weil $\eta > 0$ beliebig gewählt werden konnte, konvergiert auch

die zweite Summe des weiter oben stehenden Ausdrucks gegen 0.

Es gilt $G(\mu^n, (\tilde{\mu}_i^n)_I) \longrightarrow G(\mu, (\tilde{\mu}_i)_I)$, d.h. G ist stetig auf $A_0(\mathcal{D})$.

Q.e.d.

Bemerkungen

1. Gilt für φ nur $\limsup_{\alpha \searrow 0} \frac{\varphi(\alpha)}{\alpha} < \infty$, dann ist G bzgl. $w(x) \times \tilde{\tau}_1$

stetig. Dies kann man mit analogen Abschätzungen beweisen.

Wir werden im folgenden diese Aussage jedoch nicht benötigen.

2. Ist γ positiv und φ auf $(0,1]$ positiv, so liegt jede Clusterung mit maximaler Güte in E_I^1 . Jede zu G passende

Clusterungsfunktion hat keinen Masseverlust.

3. Ist g stetig und nicht konstant 0, so ist die Voraussetzung der Stetigkeit in Satz 1, 1) notwendig; denn wegen $g \not\equiv 0$ existiert ein $\mu \in \mathcal{D}^*$ mit $\varphi(\mu) \neq 0$ und dafür gilt $\varphi(\alpha) = \frac{g(\alpha\mu)}{\varphi(\mu)}$,

also ist φ auf $[0,1]$ notwendig stetig. Analog existiert ein $\alpha \in (0,1]$ mit $\varphi(\alpha) \neq 0$ und damit gilt für alle $\mu \in \mathcal{D}^*$ auch $\varphi(\mu) = \frac{g(\alpha\mu)}{\varphi(\alpha)}$, also ist auch φ auf $\mathcal{D}^*$ stetig.

4. Im Falle $I = \mathbb{N}$ ist die Bedingung $\lim_{\alpha \searrow 0} \frac{\varphi(\alpha)}{\alpha} = 0$ auch notwendig für die Stetigkeit von G auf $A_0(\mathcal{D})$, wenn γ nicht konstant 0 ist.

Beweis: Sei $\mu \in \mathcal{D} \setminus \{0\}$ mit $\gamma(\frac{\mu}{\mu(X)}) \neq 0$ und $(\alpha_n)_{\mathbb{N}}$ eine beliebige gegen 0 konvergente Folge in $(0,1]$. Setze

$$\mu_i^n := \begin{cases} \alpha_n \frac{\mu}{\mu(X)} & \text{für } i \leq \left\lfloor \frac{\mu(X)}{\alpha_n} \right\rfloor \\ 0 & \text{sonst.} \end{cases}$$

Es gilt $\sum_{i \in I} \mu_i^n \leq \mu$, $(\mu_i^n)_I \longrightarrow (0)_I$ bzgl. $\tilde{\tau}_\infty$ für $n \rightarrow \infty$,

also $(\tilde{\mu}_i^n)_I := \pi(\mu_i^n)_I \longrightarrow (0)_I$ bzgl. $\tilde{\tau}_\infty$ für $n \rightarrow \infty$.

Aus der Stetigkeit von G folgt für $n \rightarrow \infty$ mit $\mu \leq \mu \in \mathcal{D}^*$

$$\left[\frac{\mu(X)}{\alpha_n} \right] \cdot \varphi(\alpha_n) \cdot \gamma\left(\frac{\mu}{\mu(X)}\right) = \left[\frac{\mu(X)}{\alpha_n} \right] g(\mu_j^n) = \sum_{i \in I} g(\mu_i^n) = G(\mu, (\tilde{\mu}_i^n)_I) \longrightarrow 0, \text{ für } j \in I \text{ und genügend großes } n.$$

Wegen $\gamma(\frac{\mu}{\mu(X)}) \neq 0$ gilt also auch

$$\left[\frac{\mu(X)}{\alpha_n} \right] \cdot \alpha_n \cdot \frac{\varphi(\alpha_n)}{\alpha_n} \longrightarrow 0 \text{ für } n \rightarrow \infty. \quad \text{Da } \left[\frac{\mu(X)}{\alpha_n} \right] \alpha_n$$

wegen $\mu(X)$ konvergiert, folgt $\lim_{\alpha \searrow 0} \frac{\varphi(\alpha)}{\alpha} = 0$ für $n \rightarrow \infty$.

Da die Folge $(\alpha_n)_{\mathbb{N}}$ eine beliebige Nullfolge in $(0,1]$ war,

muß also $\lim_{\alpha \searrow 0} \frac{\varphi(\alpha)}{\alpha} = 0$ gelten.

Das Integral $\iint d(x,y) d\mu(x) d\mu(y)$ für $\mu \in \mathcal{M}^{\leq}$ ist i.a. nur dann stetig auf $\mathcal{M}^{\leq}$, wenn d stetig und beschränkt ist.

In vielen Fällen ist d aber nicht beschränkt. Im speziellen Fall $X = \mathbb{R}$ und $\mu \in \mathcal{M}^1$ (Vgl. Kendall, Stuart (1963)) nennt man das Integral für $d(x,y) = |x - y|$ die mittlere Differenz. Im Falle $d(x,y) = |x - y|^2$ gilt:

$$\iint (x-y)^2 d\mu(x) d\mu(y) = 2 \operatorname{Var}(\mu) = 2\sigma^2(\mu).$$

Im folgenden soll als wichtiges Beispiel gezeigt werden, daß für Maße mit existierendem 2. Moment die wesentliche Aussage von Satz 1 auch für $\gamma(\mu) = \frac{1}{1 + \sigma(\mu)}$ erhalten bleibt. Dazu benötigen wir einige Hilfsmittel.

Definition

Für $K \in (1, \infty)$ sei $a(K) := (-\infty, -K) \cup (K, \infty)$;

$\mathcal{V} := \{f: (1, \infty) \longrightarrow [0, \infty) : f \text{ antiton mit } \lim_{K \rightarrow \infty} f(K) = 0\}$

sei die Menge der antitonen im unendlichen verschwindenden Funktionen. Für eine stetige Funktion $g: \mathbb{R} \longrightarrow [0, \infty)$ und

$f \in \mathcal{V}$ sei:

$$\mathcal{D}_g^{\leq}(f) := \{\mu \in \mathcal{M}^1 : \int_{a(K)} g d\mu < f(K) \text{ für alle } K \in (1, \infty)\},$$

$$\mathcal{D}_g^{\leq}(f) := \{\mu \in \mathcal{M}^{\leq} : \mu \leq \nu \in \mathcal{D}_g^{\leq}(f)\}.$$

Es ist

$$\mathcal{D}_g^{\leq}(f) = \{\mu \in \mathcal{M}^{\leq} : \int_{a(K)} g d\mu < f(K) \text{ für alle } K \in (1, \infty)\}$$

Hierbei folgt die Ungleichung " $\subset$ " trivial, die Ungleichung

" $\supset$ " ergibt sich wegen $\mu \leq \mu + (1 - \mu(X)) \delta_0 \in \mathcal{M}^1$ für $\mu \in \mathcal{M}^{\leq}$.

Lemma

Sei $g: \mathbb{R} \longrightarrow [0, \infty)$ eine stetige Funktion und $f \in \mathcal{V}$.

$g^*: \mathbb{R} \longrightarrow \mathbb{R}$ sei eine beliebige stetige Funktion mit

$|g^*(x)| \leq g(x)$ für $x \geq 1$. Dann folgt aus der schwachen

Konvergenz einer Folge $(\mu^n)_{n \in \mathbb{N}}$ in $\mathcal{D}_g^{\leq}(f)$ gegen $\nu \in \mathcal{D}_g^{\leq}(f)$

auch $\int g^* d\mu^n \longrightarrow \int g^* d\mu$ für $n \longrightarrow \infty$.

Ist $\mu(X) \neq 0$, so gilt auch

$$\int g^* d \frac{\mu^n}{\mu^n(X)} \longrightarrow \int g^* d \frac{\mu}{\mu(X)} \quad \text{für } n \longrightarrow \infty.$$

Beweis:

Für $K \in (1, \infty)$ mit $\mu(\{-K, K\}) = 0$ gilt:

$$\begin{aligned} \left| \int g^* d\mu^n - \int g^* d\mu \right| &= \left| \int_{[-K, K]} g^* d\mu^n - \int_{[-K, K]} g^* d\mu + \int_{a(K)} g^* d\mu^n - \int_{a(K)} g^* d\mu \right| \\ &\leq \left| \int_{[-K, K]} g^* d\mu^n - \int_{[-K, K]} g^* d\mu \right| + \int_{a(K)} g^* d\mu^n + \int_{a(K)} g^* d\mu \leq \\ &\leq \left| \int_{[-K, K]} g^* d\mu^n - \int_{[-K, K]} g^* d\mu \right| + 2 f(K). \end{aligned}$$

Dies wird aber beliebig klein für hinreichend großes K und hinreichend großes n , da f im unendlichen verschwindet und $(\mu^n)_N$ gegen μ konvergiert mit $\mu(\{K, -K\}) = 0$. Die zweite

Behauptung folgt sofort aus der ersten.

Q.e.d.

Es ergibt sich nun der angekündigte Satz.

Satz 2

Die Funktion $g: \mathbb{R} \longrightarrow [0, \infty)$ sei durch $g(x) = x^2$ definiert und für $\mu \in \mathcal{M}^1$ existiere das 2. Moment.

1) Dann existiert eine Funktion $f \in \mathcal{V}$ mit $\mu \in \mathcal{D}_g(f)$ und mit Wahrscheinlichkeit 1 liegen schließlich alle empirischen Maße zu unabhängigen Wiederholungen einer nach μ verteilten Zufallsgröße in $\mathcal{D}_g(f)$.

2) Für jedes $f \in \mathcal{V}$ ist $\sigma: \mathcal{D}_g^{\leq}(f) \setminus \{0\} \longrightarrow [0, \infty)$ mit

$$\sigma(\mu) := \left[\int x^2 d \frac{\mu(x)}{\mu(X)} - \left(\int x d \frac{\mu(x)}{\mu(X)} \right)^2 \right]^{1/2} \quad \text{stetig und}$$

$$\gamma: \mathcal{D}_g^*(f) \longrightarrow \mathbb{R} \quad \text{mit} \quad \gamma\left(\frac{\mu}{\mu(X)}\right) = \frac{1}{1 + \sigma(\mu/\mu(X))}$$

stetig und beschränkt.

Für ein geeignetes $\mathcal{V}$ (vgl. Satz 1) ist $G: A_0(\mathcal{D}_g(f)) \longrightarrow \mathbb{R}$ definiert durch

$$* \quad G(\nu, (\tilde{\rho}_1)_I) := \sum_{1 \in I} \frac{\varphi(\rho_1(x))}{1 + \sigma(\rho_1(x)^{-1} \rho_1)} \quad (\text{mit } \sigma(\frac{0}{0})=0)$$

eine bzgl. d_A stetige Gütefunktion.

3) Ist $\mathcal{D} := \{ \mu \in \mathcal{M}^1 : \int g d\mu < \infty \}$, so ist $G: A_0(\mathcal{D}) \longrightarrow \mathbb{R}$ definiert durch * eine Gütefunktion.

Ist $(X_n)_{n \in \mathbb{N}}$ eine Folge unabhängiger Wiederholungen einer nach $\mu \in \mathcal{D}$ verteilten Zufallsgröße, so gilt mit Wahrscheinlichkeit 1 für die Folge $(\nu^n)_{n \in \mathbb{N}}$ definiert durch

$$\nu_\omega^n := \frac{1}{n} \sum_{k=1}^n \delta_{X_k(\omega)} \quad \text{für } n \in \mathbb{N}:$$

Für jedes $\varepsilon > 0$ und hinreichend großes n gilt

$$\hat{\phi}_G(\nu^n) \in U^\varepsilon(\hat{\phi}_G(\nu)) .$$

Insbesondere ist $\hat{\phi}_G$ in den Punkten $\mu \in \mathcal{D}$ konsistent, in denen G streng ist.

Beweis:

1) Es gilt:

a) Ist $\lambda \in \mathcal{M}$ ein endliches Maß auf $\mathbb{R}$, so existiert für jedes $\varepsilon > 0$ eine symmetrische, stetige, auf $[0, \infty)$ isotone Funktion $h_0: \mathbb{R} \longrightarrow [0, \infty)$ mit $h_0(x) \longrightarrow \infty$ für $|x| \longrightarrow \infty$ und $\int h_0 d\lambda < \varepsilon$.

Es genügt diese Aussage für einen Halbstrahl $[0, \infty)$ zu beweisen, d.h. zu einem endlichen Maß λ auf $[0, \infty)$ existiert zu beliebigem $\varepsilon > 0$ eine stetige, isotone Funktion $h_0: [0, \infty) \longrightarrow [0, \infty)$ mit $h_0(x) \longrightarrow \infty$ für $x \longrightarrow \infty$ und mit $\int_{[0, \infty)} h_0 d\lambda < \varepsilon$.

Da λ endlich ist, gilt $\lambda[K, \infty) \longrightarrow 0$ für $K \longrightarrow \infty$.

Zu jedem $n \in \mathbb{N}$ sei $K_n \in [0, \infty)$ gewählt mit $K_{n+1} > K_n$, $K_n \longrightarrow \infty$ und $\lambda[K_n, \infty) < \frac{1}{(n+1)^3}$. Setze $K_0 := 0$ und

$h_0^*(K_n) = n$ für $n = 0, 1, 2, \dots$ und ergänze h_0^* stückweise linear zu einer Funktion $h_0^*: [0, \infty) \longrightarrow [0, \infty)$. Dann ist h_0^* isoton und es gilt $h_0^*(x) \longrightarrow \infty$ für $x \longrightarrow \infty$ und auch:

$$\int h_0^* d\lambda = \sum_{n=1}^{\infty} \int_{[K_{n-1}, K_n)} h_0^* d\lambda \leq \sum_{n=1}^{\infty} n \lambda[K_{n-1}, K_n) \leq \sum_{n=1}^{\infty} n \lambda[K_{n-1}, \infty) \leq \sum_{n=1}^{\infty} n \frac{1}{n^3} < \infty.$$

Die Funktion $h_0 := \eta h_0^*$ erfüllt für genügend kleines $\eta > 0$ die Behauptung des Satzes für den positiven Halbstrahl $[0, \infty)$.

b) Zu $\mu \in \mathcal{D}$ existiert eine stetige, auf $[0, \infty)$ isotone, symmetrische Funktion $h: \mathbb{R} \longrightarrow [1, \infty)$ mit $h(K) \longrightarrow \infty$ für $K \longrightarrow \infty$ und $\int h g d\mu < \infty$.

Denn man kann zu dem endlichen Maß $\lambda := g\mu$ und $\varepsilon = 1$ eine Funktion h_0 nach a) wählen. Dann erfüllt $h := 1 + h_0$ diese Bedingungen.

c) Aufgrund des starken Gesetzes der großen Zahlen gilt mit Wahrscheinlichkeit 1:

$$\int h g d\mu^n \longrightarrow \int h g d\mu.$$

Also gibt es mit Wahrscheinlichkeit 1 ein $n_0 \in \mathbb{N}$, sodaß für alle $n \geq n_0$ und für $K \in (1, \infty)$ gilt:

$$\int_{a(K)} g d\mu^n \leq \frac{1}{h(K)} \int_{a(K)} h g d\mu^n \leq \frac{1}{h(K)} \int h g d\mu^n < \frac{1}{h(K)} (\int h g d\mu + 1).$$

Für f definiert durch $f(K) := \frac{1}{h(K)} (\int h g d\mu + 1)$ gilt dann $f \in \mathcal{V}$; mit Wahrscheinlichkeit 1 liegen die empirischen Maße zu μ schließlich in $\mathcal{D}_g(f)$. Es gilt 1).

2) Folgt sofort aus 1) und dem vorbereitenden Lemma.

3) Es gilt $\mathcal{D} = \bigcup_{\text{gewisse } f \in \mathcal{V}} \mathcal{D}_g(f)$ mit geeigneten $f \in \mathcal{V}$ nach

1). Also ist $G: A_0(\mathcal{D}) \longrightarrow \mathbb{R}$ eine Gütefunktion. Die weiteren Aussagen ergeben sich aus 2) in Verbindung mit

III, Satz 3.

Bock (1968,1974) hat für spezielle Verlustfunktionen aufgrund von Normalverteilungsannahmen und speziellen a priori Verteilungen für einige Parameter Risikofunktionen im Sinne von Bayes ermittelt - dies sind dann die "Gütefunktionen" -, die es zu minimieren gilt. Drei seiner "Gütefunktionen" sollen kurz im Sinne dieser Arbeit umformuliert werden. Das Varianzkriterium (Formel (11.6) in Bock (1974)) lautet nach Division durch den jeweils konstanten Stichprobenumfang für $I = \{1, 2, \dots, m\}$, $X = \mathbb{R}^p$ in der angeglichenen Formulierung mit $\sum_I \mu_i = \mu$:

$$G(\mu, (\tilde{\mu}_i)_I) := \sum_{i=1}^m \mu_i(X) \left\| \int x d \frac{\mu_i}{\mu_i(X)}(x) - \int x d \mu(x) \right\|^2.$$

Die Integrale sind hierbei natürlich komponentenweise zu verstehen. Zugelassen sind bei Bock nur empirische Maße und mengenartige, nicht leere Clusterungen. Durch eine geeignete Wahl des Definitionsbereiches $\mathcal{D} \subset \mathcal{X}^1$ wird G eine stetige Gütefunktion auf $A_0(\mathcal{D})$.

Ein weiterer Vorschlag von Bock (1974) (Formel (16.23)) lautet nach leichten Umformungen - der Einfachheit halber nur für den eindimensionalen Fall -

$$\sum_{i=1}^m \mu_i(X) \log(\text{Var}(\mu_i(X)^{-1} \mu_i)) = \text{Min} ! \quad \text{bzw.}$$

$$\sum_{i=1}^m \mu_i(X) \log(\text{Var}(\mu_i(X)^{-1} \mu_i))^{-1} = \text{Max} !$$

Man vergleiche dieses Beispiel mit Satz 1 bzw. Satz 2. Man erkennt sofort, daß diese Definition für beliebige - nicht-parametrische - Maße und Clusterungen nur dann sinnvoll sein kann, wenn μ keine Atome hat. Ein Atom μ_i würde den Ausdruck $\log(\text{Var}(\mu_i(X)^{-1} \mu_i))^{-1}$ zu ∞ entarten lassen. Als Cluster darf man also bei empirischen Maßen keine einzelnen Atome zulassen. Auf einen angemessenen Definitionsbereich auf dem die "Gütefunktion" stetig ist, soll hier nicht ein-

gegangen werden.

Als letzte Gütefunktion sei als Umformung von Formel (16.21)

$$\begin{aligned} G(\mu, (\tilde{\mu}_i)_I) &= - \sum_{i=1}^m \int \|x - y\| d \frac{\mu_i}{\mu_i(\bar{X})}(y) \|^2 d\mu_i(x) = \\ &= - \frac{1}{2} \sum_{i=1}^m \iint \|x - y\|^2 d\mu_i(x) d\mu_i(y) \end{aligned}$$

angegeben. Man vergleiche hierzu die Bemerkungen oben zu verallgemeinerten Streuungsmaßen.

Insgesamt zeigt Bock, daß man aufgrund bestimmter Modellvorstellungen zu sinnvollen Gütefunktionen gelangen kann. Es wird allerdings in vielen Fällen notwendig sein, diese Funktionen im Sinne der robusten Statistik zu modifizieren. Weiterhin sind bei allen diesen Gütefunktionen die Anzahl der nicht leeren Cluster vorgeschrieben. Dies liegt im wesentlichen an der Ordnung 1 des Faktors $\mu_i(X)$ in den Definitionen der Gütefunktionen.

Gordon und Henderson (1977) schlagen einen Algorithmus (nicht-lineares Programmieren) für das oben aufgeführte Varianzkriterium vor, vgl. Bock (1974) Formel (11.7).

In ihrem Lemma weisen sie nach, daß die zugehörige verallgemeinerte Clusterungsfunktion auf empirischen

Maßen mengenartig ist, d.h. für $(\tilde{\mu}_i)_I \in \hat{\phi}_G(\mu)$ mit $\mu = \frac{1}{n} \sum_{i=1}^n \delta_{x_i}$ folgt, $(\tilde{\mu}_i)_I$ ist mengenartig.

Anhang

Zu den Überlegungen des Kapitels III gibt es in der allgemeinen Schätztheorie Parallelen. Ist $\mathcal{D} \subset \mathcal{M}^1(\mathbb{R}^P)$ eine Menge von W-Maßen, Θ eine Parametermenge und $G: \mathcal{D} \times \Theta \longrightarrow \mathbb{R}$ eine Funktion, so kann man versuchen eine "Schätzfunktion" $\phi_G: \mathcal{D} \longrightarrow \Theta$ durch $G(\mu, \phi_G(\mu)) = \sup_{\vartheta \in \Theta} G(\mu, \vartheta)$ implizit zu definieren; vgl. Huber's (1966) M-Schätzer.

Definition

Zu zwei nicht leeren Mengen $\mathcal{D}$ und Θ sei eine Funktion $G: \mathcal{D} \times \Theta \longrightarrow \mathbb{R}$ gegeben. Die Abbildung

$M_G: \mathcal{D} \longrightarrow \mathbb{R}$ sei durch $M_G(\mu) := \sup_{\vartheta \in \Theta} G(\mu, \vartheta)$ definiert.

Weiter sei

$$\hat{\phi}_G(\mu) := \{\vartheta_0 \in \Theta : G(\mu, \vartheta_0) = M_G(\mu)\}.$$

Wir sagen G ist streng, wenn $|\hat{\phi}_G(\mu)| \leq 1$ für alle $\mu \in \mathcal{D}$ gilt.

Es gilt nun der

Satz

$\mathcal{D}$ und Θ seien metrisierbare, topologische Räume und Θ sei kompakt.

a) Zu einer stetigen Abbildung $\phi: \mathcal{D} \longrightarrow \Theta$ gibt es eine strenge, stetige, reelle Funktion $G: \mathcal{D} \times \Theta \longrightarrow \mathbb{R}$, die ϕ implizit definiert.

b) Eine strenge, stetige, reelle Funktion $G: \mathcal{D} \times \Theta \longrightarrow \mathbb{R}$ definiert implizit eine stetige Funktion $\phi: \mathcal{D} \longrightarrow \Theta$.

Beweis: a)

Sei d eine Metrik auf $\mathcal{D} \times \Theta$, die die Produkttopologie induziert.

$A := \{(\mu, \phi(\mu)) : \mu \in \mathcal{D}\}$ ist als Graph der stetigen Funktion ϕ abgeschlossen in $\mathcal{D} \times \Theta$. Für die Funktion $G_\phi: \mathcal{D} \times \Theta \longrightarrow \mathbb{R}$ definiert durch $G_\phi(\mu, \vartheta) := -d((\mu, \vartheta), A)$ gilt die Aussage des Satzes. a).

b) 1. Da Θ kompakt ist, wird das Supremum angenommen.

2. M_G ist nach unten halbstetig.

3. Da G streng ist, ist ϕ eindeutig definiert.

4. Es genügt die Folgenstetigkeit von ϕ nachzuweisen. Für einen indirekten Beweis nehmen wir an: Die Folge $(\mu^n)_{n \in \mathbb{N}}$ konvergiere in $\mathcal{D}$ gegen $\mu \in \mathcal{D}$, aber für ein $\varepsilon > 0$ und jedes $n \in \mathbb{N}$ gäbe es ein $n' \geq n$ mit $\phi(\mu^{n'}) \notin U^\varepsilon(\phi(\mu))$. Da $\mathcal{C}$ folgenkompakt ist, existiert eine Teilfolge $(\phi(\mu^{n'}))_{n' \in \mathbb{N}}$ von $(\phi(\mu^{n'}))_{n' \in \mathbb{N}}$ mit $\phi(\mu^{n'}) \rightarrow \mathcal{J}_0 \in \mathcal{C}$. Da G stetig ist, gilt

$$M_G(\mu^{n'}) = G(\mu^{n'}, \phi(\mu^{n'})) \rightarrow G(\mu, \mathcal{J}_0) .$$

Wegen 2. also:

$$\liminf M_G(\mu^{n'}) = G(\mu, \mathcal{J}_0) \geq M_G(\mu) ,$$

also $\mathcal{J}_0 = \phi(\mu)$, im Widerspruch zur Annahme.

Q.e.d.

Bemerkung:

M_G ist stetig.

Die folgende Charakterisierung robuster Schätzungen ergibt sich sofort aus dem Satze.

Ist $\mathcal{C}$ eine kompakte, metrisierbare Menge von Parametern, (X, d) ein separabler, metrischer Raum, $\mathcal{D}$ eine Menge von W -Maßen, die mit Wahrscheinlichkeit 1 zu jedem $\mu \in \mathcal{D}$ schließlich alle empirischen Maße zu μ enthält: Dann sind die robusten Schätzer auf $\mathcal{D}$ gerade die Schätzer, die durch strenge, stetige, reelle Funktionen G implizit definiert werden.

- Balachandran, V.K. A mapping theorem for metric spaces.
Duke Mathematical Journal 22 (1955)
- Billingsley, P. Convergence of probability measures.
New York, Wiley 1968
- Bock, H.H. Statistische Modelle für die einfache und doppelte Klassifikation von normalverteilten Beobachtungen. Dissertation
Universität Freiburg i. Brsg. 1968
- Automatische Klassifikation
Vandenhoeck und Ruprecht, Göttingen 1974
- Bourbaki, N. General Topology I
Paris, Hermann 1966
- Gordon, A.D. An algorithm for Euclidean sum of squares classification.
Henderson, J.T. Biometrics 33 (1977) 355-362
- Hanisch, W. Clusterungen. Diplomarbeit (bei U. Oppel)
am Mathematischen Institut der Universität München, 1977
- Hartigan, J.A. Clustering Algorithms
New York, Wiley 1975
- Huber, P.J. Applications vs. abstraction: The selling out of mathematical statistics?
Proceedings of the Conference on Directions for Mathematical Statistics.
Suppl. adv. Appl. Prob. 7, (1975 a) 84-89
- Some Mathematical Problems Arising in Robust Statistics.
Proceedings of the International Congress of Mathematicians Vancouver 1974 , (1975 b) 153-157
- The Behavior of Maximum Likelihood Estimates under Nonstandard Conditions.
Proceedings of the fifth Berkeley Symposium on Mathematical Statistics and Probability. 1966
- Jardine, N. Mathematical Taxonomy
Sibson, R. New York - London, Wiley 1971
- Kellerer, H.G. Markov-Komposition und eine Anwendung auf Martingale.
Math. Ann. 198 (1972) 99-122
- Kendall, M.G. The advanced theory of statistics I
Stuart, A. London, Charles Griffin Company 1963
- Ling, R.F. A Probability Theory of Cluster Analysis
JASA 68 (1973) 159-164
- Varadarajan, V.S. Measures on topological spaces
Am. Math. Soc. Translation Series 2, 48 (1966) 161-228
- Rao, C.R. Recent trends of research work in multivariate analysis, Biometrics 28 (1972) 3-22

Lebenslauf:

Geburtsdatum: 10. 4. 1946

Geburtsort: Köln

Schulbildung: Nach vierjähriger Volksschulzeit besuchte ich das altsprachliche St. Michael Gymnasium in Bad Münster-eifel und bestand dort das Abitur im Frühjahr 1965.

Studium: Mathematik:
Universität Köln, SS 65 - WS 68/69.
Insbesondere bei den Professoren Dombrowski, Hajdu, Jehne, Schäfke, Schröder.
Universität München, SS 69 - WS 70/71
Insbesondere bei den Professoren Diller, Hämmerlin, Kasch, Mammitzsch, Richter, Roelcke, Schütte.

Medizin:
Universität München, seit 1971.

Abschlußprüfungen: Die Diplomhauptprüfung legte ich im März 1971 in München ab.

Berufliche Tätigkeiten: Vom Sommer 1971 bis Dezember 1976 war ich am Max-Planck-Institut für Psychiatrie in der Abteilung für Biostatistik zuerst als Stipendiat, dann als wissenschaftlicher Assistent tätig.



difo
druck

laubanger 15
8600 bamberg
tel.: 09 51/6 43 21